

**Epigenetic control of endogenous retroviruses
and their immune recognition in differentiated
human cells**

Christopher Hieng Chie Tie

Thesis submitted for the degree of Doctorate of

Philosophy

University College London

September 2017

I, Christopher Hieng Chie Tie, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

First and foremost, I would like to thank my primary supervisor, Helen, for your supervision, patience, and encouragement throughout my journey. I would also like to express my gratitude towards UCL for supporting me financially.

Next, I thank Greg for your critical discussion on my project and introducing me to 'Watch Me'. I am also grateful for my thesis committee, Benny and Yasu, for your support and guidance. I also thank Luisa, Petra, and MT for your help and technical knowledge on experimental design. Massive thanks to Jane T for keeping the lab running and putting up with all the lab shenanigans. I thank Carolina, Claire, Rich, Stephen, Lauren, Lorena, and Tom for being my personal peer support group and special shout out to Lauren for providing the party house. I appreciate all the help and banter from everyone in wing 3.3 for making my experience an amazing one!

I thank all my friends for your patience and understanding throughout the last 4 years. I would like to extend my appreciations to Sam, Alice, Michelle, Cherie, Terrence and CY Lim for being there whenever I need someone to have a meal or a glass of Negroni with. Special thanks to Sam for being understanding and a great housemate (most of the time).

For my families, I thank you for your continuous love and support. I promise one day I would explain what I actually work on without sounding like I am speaking Martian.

Finally, I would also like to dedicate my work to my late grandparents who missed the chance to witness my achievement but your encouragement and reminders will always be with me.

Abstract

Endogenous retroviruses (ERVs) have accumulated in vertebrate genomes and contribute to the complexity of gene regulation. ERVs are beneficial to their hosts because their promoters can drive temporal and spatial expression of cellular genes. Aberrantly activated ERVs, however, can be detrimental. ERV transcription is therefore controlled by multiple epigenetic modifiers. One ERV repression pathway involves KAP1 and Krüppel-associated box domain-zinc finger proteins (KZNFs), which recruit KAP1 to ERVs and other repetitive sequences early in development. Little is known, in contrast, about the regulation of ERVs in differentiated cells, particularly in humans. In this thesis, we aimed firstly to address the question of whether KAP1 and related epigenetic factors are necessary to repress ERVs in differentiated human cells. Secondly, we sought to assess the impact of ERV reactivation on the innate immune system. We found through KAP1 knockout and mRNA-sequencing analyses that KAP1 represses ERVs and ZNFs, both of which overlap with KAP1 binding sites and silent chromatin marks in multiple cell types. Furthermore, this pathway is functionally conserved in primary human peripheral blood mononuclear cells (PBMCs). We show that cytosine methylation that acts on KAP1-regulated loci is necessary to prevent immune reactivity of ERVs and other retrotransposons that can mimic viruses by producing immunostimulatory nucleic acids. While KAP1 depletion alone leads to activation of several chemokines, it is not sufficient for global induction of interferon-stimulated genes. Therefore, we depleted key epigenetic complexes that KAP1 collaborates with including the HUSH complex comprising MPP8, Periphilin and Tasor. We identified MPP8 to play a dominant role in preventing aberrant immune activation in human cells. MPP8 is a chromodomain protein implicated in the spread of heterochromatin. In sum, these data indicate that the KAP1-KZNF pathway and MPP8 are central to genome stability and the control of viral mimicry in differentiated human cells.

Abbreviations

5-Aza	5-Aza-2'-deoxycytidine
Ago	Argonaute
AIM2	Absent In Melanoma 2
ALS	Amyotrophic Lateral Sclerosis
Alu	Arthrobacter luteus
ALV	Avian Leucosis Virus
ATRX	α -Thalassemia / mental Retardation syndrome X-linked
B2M	Beta-2-Microglobulin
CD	Cluster of Differentiation
cGAMP	2'3' cyclic-AMP-GMP
cGAS	cyclic GMP-AMP Synthase
DAI	DNA-dependent Activator of IFN regulatory factor
DAMP	Danger-Associated Molecular Pattern
DAXX	Death-domain-Associated protein
DLBCL	Diffuse Large B Cell Lymphoma
DNMT	DNA Methyltransferase
dNTP	deoxynucleotide
dsRNA	double stranded RNA
EBP1	ErbB3-Binding Protein 1
ENV	Envelope
eRNA	enhancer RNA
ERV	Endogenous Retrovirus
ESC	Embryonic Stem Cell
ETn	Early Transposable element
GAG	Group-specific Antigen
H3K27ac	acetylation of Histone 3 at Lysine 27
H3K27me3	trimethylation of Histone 3 at Lysine 27
H3K4me1	monomethylation of Histone 3 at Lysine 4
H3K9me3	trimethylation of Histone 3 at Lysine 9
H4K20me3	trimethylation of Histone 4 at Lysine 20

HDAC	Histone Deacetylase
HEK293	Human Embryonic Kidney 293 cells
HEK293T	Human Embryonic Kidney 293 cells with T-antigen
HERV	Human Endogenous Retrovirus
HIV	Human Immunodeficiency Virus
HL	Hodgkin's Lymphoma
HMTase	Histone Methyltransferase
HP1	Heterochromatin Protein 1
HSP90	Heat Shock Protein 90
HUSH	Human Silencing Hub
IAP	Intracisternal A Particle
IDDM	Insulin-Dependent Diabetes Mellitus
IFI16	IFN- γ -inducible protein 16
IFN	Interferon
IL	Interleukin
iPSC	induced Pluripotent Stem Cell
IRF5	Interferon Regulatory Factor 5
ISG	Interferon-stimulated gene
ISRE	IFN-Stimulated Regulatory Element
JSRV	Jaagsiekte Sheep Retrovirus
KAP1	KRAB-Associated Protein 1
KRAB	Krüppel-associated box
kb	kilo base pairs
KD	Knockdown
KO	Knockout
KZNF	KRAB Zinc Finger Protein
LINE	Long Interspersed Nuclear Element
lncRNA	long non-coding RNA
LSD1	Lysine Specific Demethylase 1A
LSH1	Lymphoid-Specific Helicase 1
LTR	Long Terminal Repeat

LTR	Long Terminal Repeat
MaLR	Mammalian apparent LTR Retrotransposon
MDA5	Melanoma Differentiation-Associated gene 5
MeCP2	Methyl-CpG-binding Protein 2
MERV-L	Murine ERV-L
miRNA	microRNA
MLV	Murine Leukemia Virus
MMTV	Mouse Mammary Tumour Virus
MPP8	MPHOSPH8
mRNA	messenger RNA
MS	Multiple Sclerosis
MyD88	Myeloid Differentiation factor 88
NPC	Neural Progenitor Cells
ORF	Open Reading Frame
ORF1p	Open Reading Frame I product
ORF2p	Open Reading Frame II product
PAMP	Pathogen-Associated Molecular Patterns
PBMC	Peripheral Blood Mononuclear Cell
PBS	Primer Binding Site
PHD	Plant Homeodomain
piRNA	Piwi-interacting RNA
POL	Polymerase
Pol II	RNA Polymerase II
Pol III	RNA Polymerase III
Poly-A	Polyadenylation
pri-miRNA	primary miRNA transcript
PRR	Pattern Recognition Receptor
RA	Rheumatoid Arthritis
RBCC	RING-B-box-Coiled-coil
RIG-I	Retinoic acid-Inducible Gene I
RISC	RNA-Induced Silencing Complex

RLP	Retroviral Like Particle
RLR	RIG-I Like Receptor
RNP	Ribonucleoprotein
RT	Reverse transcriptase
SINE	Short Interspersed Nuclear Element
siRNA	short / small interfering RNA
SLE	Systemic Lupus Erythematosus
ssDNA	single stranded DNA
STING	Stimulator of IFN genes
SVA	SINE-VNTR- <i>A/u</i>
TE	Transposable Element
TFIIB	Transcription Factor II B
TFIIC	Transcription Factor II C
TFRC	Transferrin Receptor
Th1	Type 1 T helper cell
TIR	Toll-IL-1-Resistance
TLR	Toll-Like Receptor
TM	Transmembrane
TNF	Tumour Necrosis Factor
TPRT	Target site Primed Reverse Transcription
tRNA	transfer RNA
TSA	Trichostatin A
TSD	Target Site Duplication
UTR	Untranslated Region
VLP	Virus Like Particle
VSV	Vesicular stomatitis virus
VNTR	Variable Number of Tandem Repeat
XRV	Exogenous Retrovirus
ZFP	Zinc Finger Protein
ZNF	Zinc Finger Protein

Table of contents

Chapter 1: Introduction	14
1.1 What are transposable elements (TEs)?.....	14
1.1.1 ERVs	17
1.1.1.1 Life cycle of ERVs.....	19
1.1.1.2 Classification.....	19
1.1.2 LINEs	21
1.1.2.1 Life cycle of LINE1.....	23
1.1.2.2 Classification.....	25
1.1.3 SINEs	27
1.1.3.1 Life cycle of SINEs	27
1.1.3.2 SVAs.....	31
1.2 The impact of ERVs on their hosts	31
1.2.1 Exaptation of ERVs in host gene regulation	31
ERVs as alternative promoters.....	33
ERVs as lncRNAs	34
ERVs as enhancers	35
ERVs as pathogen-associated molecular patterns (PAMPs).....	36
1.2.2 ERVs and cellular protein expression.....	36
Novel cellular protein expression patterns.....	36
ERVs as antiviral agents	38
1.2.3 Negative effects of ERVs	39
1.3 Regulation of ERVs	44
1.3.1 Epigenetics	44
1.3.2 Epigenetic regulation of ERVs in early development.....	46
KZNFs and KAP1	46
YY1 and EBP1	49
Heat Shock Protein 90	50
Histone modifications.....	50
H3.3 and HUSH	52
Histone deacetylation.....	53
DNA methylation	53
1.3.3 ERV regulation in somatic cells	55
1.3.4 Small RNA-mediated repression.....	56
1.3.5 Restriction factors.....	57
1.4 ERVs and the immune system	59
1.4.1 Toll-like receptors.....	60
1.4.2 RNA sensing.....	63
1.4.3 DNA sensing.....	65
1.4.4 Consequences of innate immune activation	67
1.4.5 ERVs and adaptive immunity	68
Rationale.....	71
Aims.....	73
Aim 1. KAP1 repression	73
Aim 2. Biological relevance.....	74
Chapter 2: Materials and Methods	75
2.1 General solutions, buffers and materials	75
2.2 Western Blotting.....	75
2.2.1 SDS polyacrylamide gel electrophoresis (PAGE).....	75
2.2.2 SDS-PAGE running	76
2.2.3 Transfer	76
2.2.4 Protein visualisation	76
2.2.5 Antibodies.....	77

2.3 Cell culture	77
2.3.1 Antibiotics for cell culture	78
2.3.2 Cell culture maintenance	78
2.3.3 Primary cell isolation	79
2.4 CRISPR/Cas9 genome editing	79
2.5 Cloning	80
2.5.1 DNA preparation	80
2.5.2 Bacterial Transformation	80
2.5.3 Oligo annealing for shRNA cloning	81
2.5.4 Ligation	81
2.5.5 Sequencing	82
2.6 Lentiviral vectors	82
2.6.1 Vector Production	82
2.6.2 Virus titration	82
2.7 Polymerase chain reaction (PCR)	83
2.7.1 PCR for Bisulfite sequencing	83
2.7.2 Phusion PCR	84
2.8 HERVK14C LTR reporter	86
2.8.1 Cloning	86
2.8.2 Reporter experimental set-up	86
2.9 Flow cytometry	86
2.9.1 Intracellular staining	87
2.9.2 Buffers	87
2.10 Luciferase assay	88
2.10.1 Transfection	88
2.10.2 Luciferase readout	88
2.11 RNA extraction	89
2.12 Quantitative reverse transcription polymerase chain reaction (qRT-PCR)	89
2.12.1 cDNA synthesis	89
2.12.2 Reaction mix	90
2.12.3 Cycling protocol	91
2.13 Taqman qPCR	91
2.14 RNA sequencing	92
2.14.1 Library preparation	92
2.14.2 Sequencing	93
2.14.3 Data analysis (Genes)	93
2.14.4 Data analysis (Repeats)	93
2.15 Chip-sequencing analysis	95
2.16 DNA extraction	95
2.17 DNA methylation analysis	95
2.18 Statistical analysis	96
2.19 Oligonucleotides	96
2.20 Plasmid list	103
2.21 Drug list	104
2.22 Data access	104
Chapter 3: KAP1 regulates ERVs in differentiated human cells	105
3.1 HERVK14C as a model to study the role of KAP1 in differentiated cells	105
3.2 KAP1 mediates HERVK14C repression in undifferentiated NTERA-2 cells	108
3.3 KAP1 represses HERVK14C in differentiated cells	112
3.4 KAP1 binds to and represses cellular genes	117
3.5 KAP1 represses multiple HERV families in differentiated cells	117

3.6 KAP1 knockout leads to downregulation of certain cellular genes and repeats	118
3.7 KAP1 binds to common ERVs and ZNFs in undifferentiated and differentiated cells	120
3.8 KAP1 common binding sites are enriched for repressive chromatin marks ...	122
3.9 KAP1 mediates de novo repression of incoming ERVs through the primer binding site sequence.....	124
Discussion I	128
Chapter 4: RKAP1 regulates ERVs in primary human cells	131
4.1 KAP1 knockdown in PBMCs leads to upregulation of retrotransposons	131
4.2 KAP1 regulates multiple KZNFs in PBMCs.....	131
4.3 KAP1 knockdown is not sufficient for robust upregulation of HERVK14C in CD4+ T cells.....	134
4.4 KAP1 can restrict incoming ERVs through the PBS sequence in PBMCs	134
4.5 Multiple KZNFs are widely expressed in various differentiated cell types.....	136
Discussion II	139
Chapter 5: Epigenetic control of ERVs prevents viral mimicry	142
5.1 DNA methylation and histone acetylation regulate retrotransposons.....	142
5.2 KAP1 contributes to the prevention of an aberrant innate immune response	145
5.3 5-Aza treatment triggers an ISG response	148
5.4 There is redundancy between KAP1 repression and DNA methylation	151
5.5 KAP1 is involved in the maintenance of DNA methylation at ERVs.....	151
5.6 The role of epigenetic regulators that KAP1 collaborates with in the prevention of viral mimicry	154
5.7 MPP8 protects cells from aberrant ISG activation.....	157
5.8 MPP8 depletion activates classical type I IFN signalling involving IRF3 and JAK-STAT	162
Discussion III.....	165
Chapter 6: Epilogue and future directions	170
6.1 Significance of this thesis.....	170
6.2 Future directions	171
References.....	173
Appendix A.....	202
Appendix B.....	203

Table list

TABLE 1 LIST OF COMMONLY USED MATERIALS AND BUFFERS.	75
TABLE 2 COMPOSITION OF SDS-PAGE USED IN THE STUDIES.	75
TABLE 3 ANTIBODIES USED IN THE STUDIES.	77
TABLE 4 CELL TYPES USED THROUGHOUT THE STUDIES.	77
TABLE 5 ANTIBODIES USED FOR CELL SELECTION IN THE STUDIES.	78
TABLE 6 ANTIBIOTIC USED FOR BACTERIAL SELECTION IN THE STUDIES.	81
TABLE 7 REACTION MIX FOR HOT START PCR.	84
TABLE 8 CYCLING PROTOCOL FOR HOT START PCR.	84
TABLE 9 REACTION MIX FOR PHUSION PCR.	85
TABLE 10 CYCLING PROTOCOL FOR PHUSION PCR.	85
TABLE 11. BUFFERS USED FOR INTRACELLUALR STAINING.	87
TABLE 12 TRANSFECTION MIX FOR LUCIFERASE BASED ASSAY.	88
TABLE 13 REACTION MIX FOR QRT-PCR.	90
TABLE 14 REAGENTS USED IN QRT-PCR REACTION.	90
TABLE 15 CYCLING PROTOCOL USED FOR QRT-PCR.	91
TABLE 16 REACTION MIX USED FOR TAQMAN QPCR.	91
TABLE 17 REAGENTS USED IN TAQMAN QPCR REACTION.	91
TABLE 18 CYCLING PROTOCOL USED FOR TAQMAN QPCR.	92
TABLE 19 PRIMERS FOR QRT-PCR IN THE STUDIES.	96
TABLE 20 shRNAs FOR KNOCKDOWN IN THE STUDIES.	99
TABLE 21 sgRNAs FOR GENERATING CRISPR KNOCKOUT CELLS.	101
TABLE 22 PCR PRIMERS USED FOR BISULFITE SEQUENCING.	101
TABLE 23 SEQUENCING PRIMERS USED THROUGHOUT THE STUDIES.	102
TABLE 24 PRIMERS AND PROBES USED FOR TAQMAN QPCR.	102
TABLE 25 PRIMERS USED FOR AMPLIFYING GENOMIC SEQUENCE IN REPORTER PLASMID GENERATION.	102
TABLE 26 LIST OF PLASMIDS USED IN THE STUDIES.	103
TABLE 27 LIST OF DRUGS USED IN THE STUDIES.	104

Figure list

FIGURE 1 TYPES OF TRANSPOSABLE ELEMENTS (TEs) IN THE HUMAN GENOME.	15
FIGURE 2 STRUCTURE OF ENDOGENOUS RETROVIRUSES (ERVs).	16
FIGURE 3 LIFE CYCLE OF ERVs.	18
FIGURE 4 STRUCTURE OF A LINE1 (L1) ELEMENT.	22
FIGURE 5 L1 RETROTRANSPOSITION MECHANISM.	24
FIGURE 6 EVOLUTION OF L1.	26
FIGURE 7 STRUCTURE OF <i>ALU</i> AND <i>SVA</i> ELEMENTS.	29
FIGURE 8 L1-MEDIATED SINE RETROTRANSPOSITION MECHANISM.	30
FIGURE 9 MECHANISMS OF ERV CO-OPTION.	32
FIGURE 10 PUTATIVE EPIGENETIC MODIFIERS INVOLVED IN ERV REPRESSION IN DIFFERENTIATED CELLS.	45
FIGURE 11 TLR SIGNALLING IN RESPONSE TO LIGAND BINDING.	62
FIGURE 12 CYTOSOLIC PATHOGEN-RECOGNITION RECEPTOR (PRR) SIGNALLING.	66
FIGURE 13 HERVK14C AS A MODEL TO EXPLORE THE ROLE OF KAP1.	106
FIGURE 14 DESIGN OF NEW HERVK14C PRIMER SETS.	107
FIGURE 15 KAP1 REGULATES HERVK14C IN NTERA-2 CELLS.	109
FIGURE 16 KAP1 REPRESSES HERVK14C IN DIFFERENTIATED CELLS.	110
FIGURE 17 ANALYSIS OF KAP1 REPRESSED CELLULAR GENES.	114
FIGURE 18 ANALYSIS OF HERVs REPRESSED BY KAP1.	116
FIGURE 19 ANALYSIS OF CELLULAR GENES DOWNREGULATED UPON KAP1 KO.	119
FIGURE 20 ANALYSIS OF REPEATS DOWNREGULATED UPON KAP1 KO.	119
FIGURE 21 ANALYSIS OF CONSERVED KAP1 BINDING LOCI AS REVEALED THROUGH CHIP-SEQ.	121
FIGURE 22 KAP1 COMMON BINDING SITES ARE ENRICHED FOR REPRESSIVE HISTONE MARKS AND OVERLAP WITH KZNF-BOUND ERV LOCI.	123
FIGURE 23 KAP1 IS CRUCIAL FOR PBS-MEDIATED DE NOVO SILENCING OF INCOMING ERVs.	125
FIGURE 24 MODELS.	127
FIGURE 25 KAP1 REPRESSION OF ERVs IS CONSERVED IN PBMCS.	132
FIGURE 26 KAP1 DEPLETION LEADS TO LITTLE OR NO HERVK14C UPREGULATION IN CD4+ T CELLS.	135
FIGURE 27 PBS-MEDIATED DE NOVO REPRESSION OF ERVs IS CONSERVED IN PRIMARY CELLS.	137
FIGURE 28 ZNFs ARE EXPRESSED WIDELY AT THE mRNA LEVEL.	138
FIGURE 29 MODEL.	140
FIGURE 30 KAP1-REGULATED RETROTRANSPOSONS ARE SUBJECT TO DNA METHYLATION.	143
FIGURE 31 KAP1 DEPLETION AFFECTS EXPRESSION OF ISGs.	146
FIGURE 32 5-AZA TREATMENT LEADS TO AN UPREGULATION OF KAP1-REGULATED RETROTRANSPOSONS AND ROBUST ISG INDUCTION IN CANCER CELL LINES AND PRIMARY CELLS.	149
FIGURE 33 COMBINING KAP1 KNOCKDOWN WITH 5-AZA TREATMENT DID NOT HAVE AN ADDITIVE EFFECT.	152
FIGURE 34 KAP1 IS REQUIRED TO MAINTAIN DNA METHYLATION AT HERVK14C.	155
FIGURE 35 MULTIPLE EPIGENETIC MODIFIERS ARE INVOLVED IN REPRESSING ERVs.	156
FIGURE 36 MPP8 PREVENTS VIRAL MIMICRY AND INNATE IMMUNE ACTIVATION.	158
FIGURE 37 MPP8 DEPLETION LEADS TO A STRIKING INDUCTION OF ISGs WHILE KAP1-REGULATED RETROTRANSPOSONS ARE DOWNREGULATED.	160
FIGURE 38 MPP8 DEPLETION LEADS TO ACTIVATION OF AN IFN β REPORTER.	163
FIGURE 39 MPP8 DEPLETION LEADS TO ACTIVATION OF IRF3 AND JAK-STAT DEPENDENT REPORTERS.	164
FIGURE 40 EPIGENETIC REGULATORS PROTECT CELLS FROM AN UNWANTED INNATE IMMUNE RESPONSE.	166

Chapter 1: Introduction

1.1 What are transposable elements (TEs)?

When Barbara McClintock first discovered a ‘controlling element’ in the maize genome in the 1950s, the idea of transposon-mediated gene regulation was novel and controversial (McClintock, 1956). This new paradigm was further developed by Britten and Davidson and they subsequently postulated a model in which repeat families could amplify a set of regulatory elements resulting in complex gene regulatory networks (Britten and Davidson, 1971). Despite so, repetitive DNA is still often referred to as “selfish-DNA” or “junk DNA”. TEs remain understudied (Orgel and Crick, 1980) but with the advancement of genome sequencing and bioinformatics tools, scientists are starting to comprehend the proportion of these ‘controlling elements’ in mammalian genomes (Lander et al., 2001). Their identities and functions are also starting to be revealed.

TEs refer to a stretch of DNA sequence that is mobile. TEs can be divided into two classes namely retrotransposons and DNA transposons. Retrotransposons are capable of replicating themselves via the copy-and-paste mechanism through a RNA intermediate reminiscent to the replication mechanism used by retroviruses like HIV. Retrotransposons are successful in colonising our genome and they have accumulated to at least 50% of the human genome (de Koning et al., 2011; Goodier and Kazazian, 2008; Rebollo et al., 2012). Retrotransposons can be further divided into long terminal repeats (LTR) retrotransposons and non-LTR retrotransposons (Figure 1).

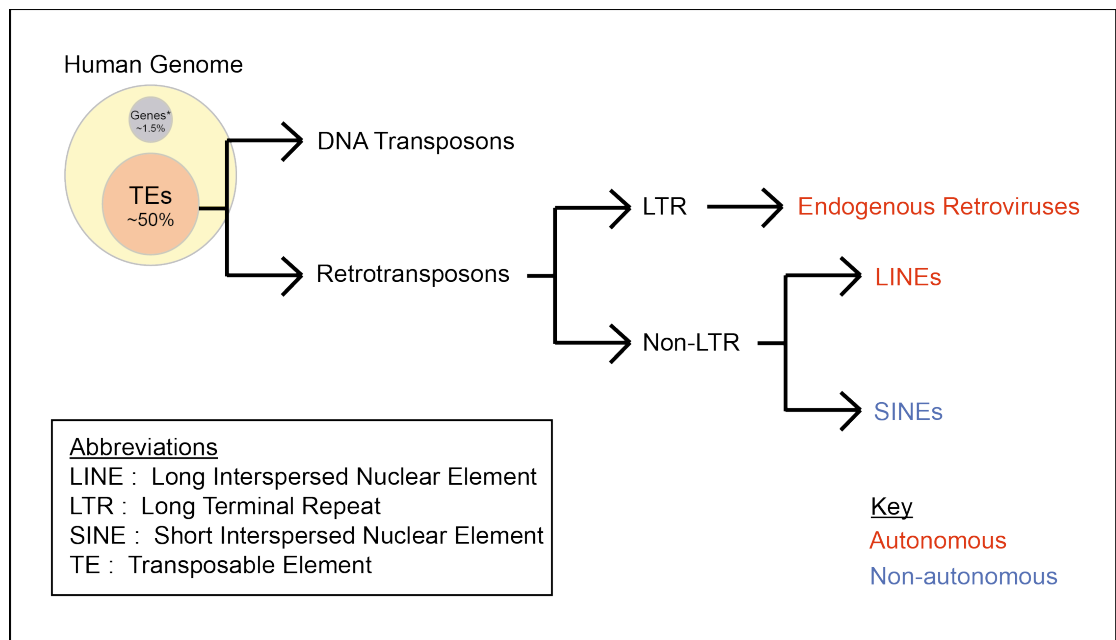


Figure 1 Types of transposable elements (TEs) in the human genome.

The majority of transposon-derived repeats in the human genome are retrotransposons, which represent 50% of our genome. TEs can be categorised based on the presence of LTRs and on whether or not they can replicate autonomously. ERVs and LINEs are autonomous families while SINEs are non-autonomous.

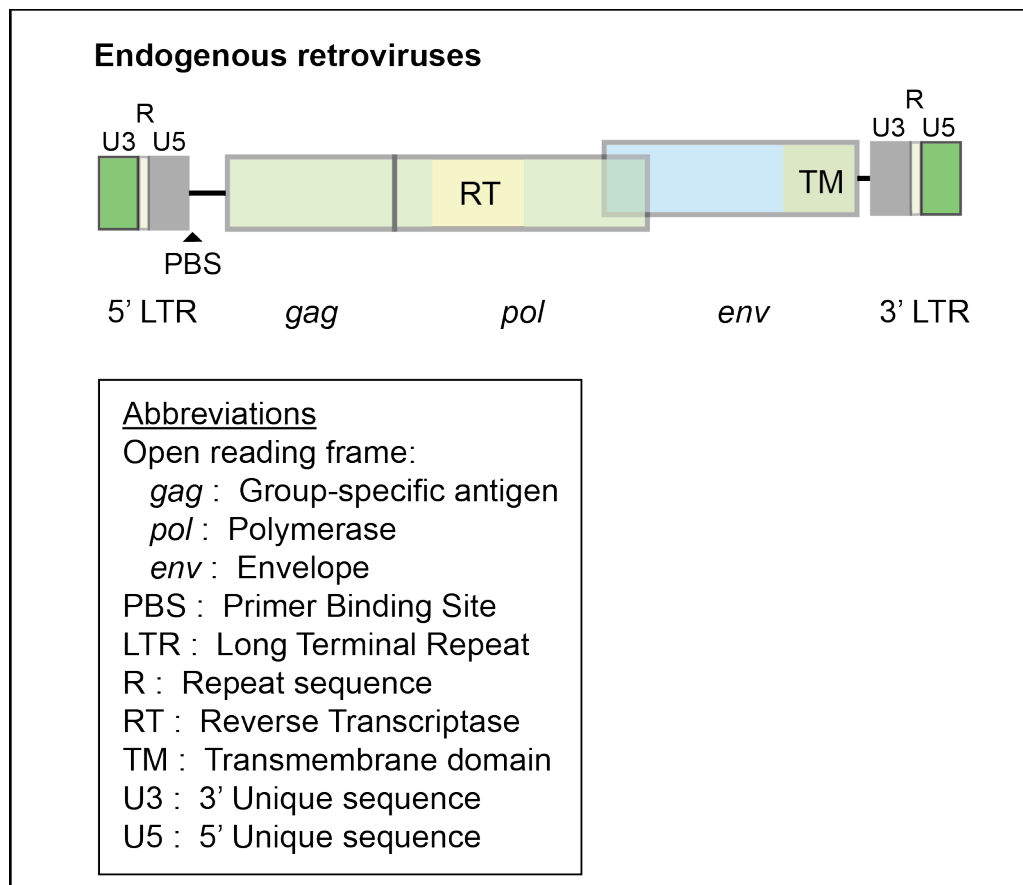


Figure 2 Structure of endogenous retroviruses (ERVs).

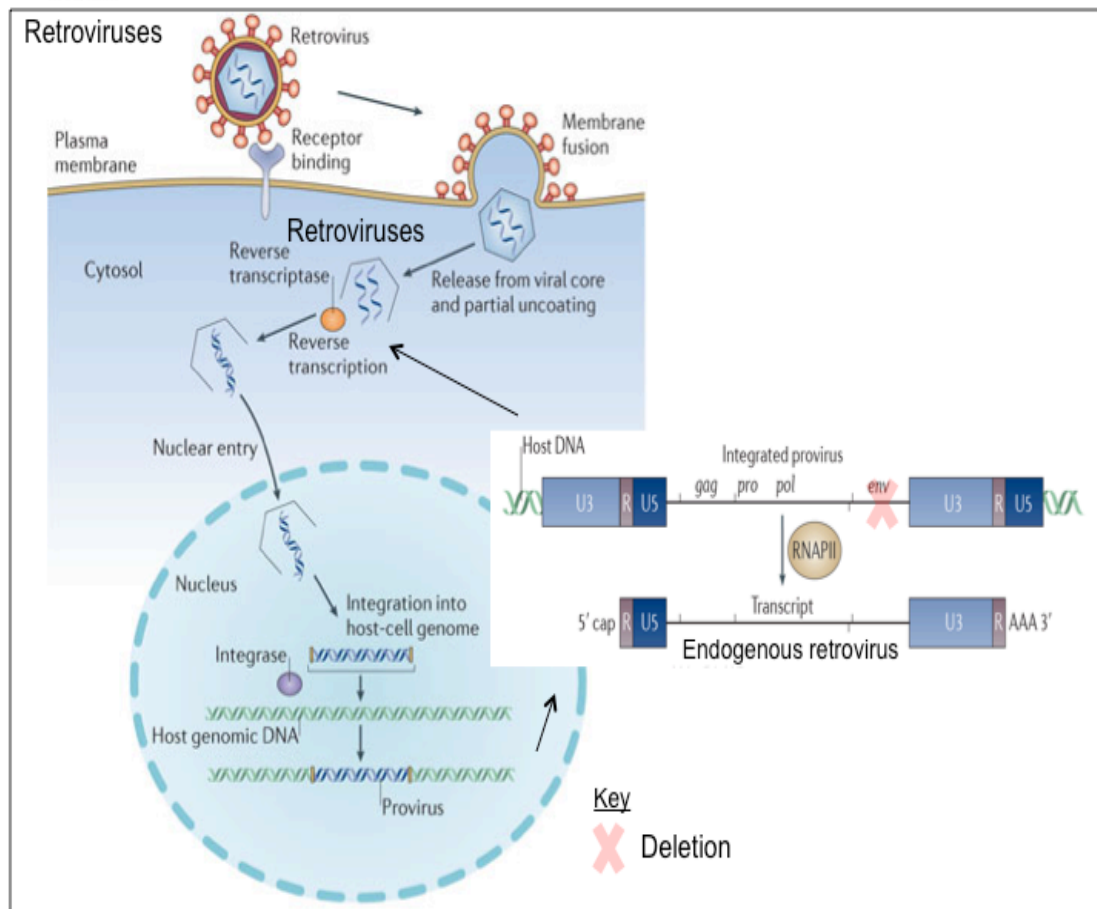
A canonical ERV provirus with *gag*, *pol*, and *env* genes is shown here. The identical long-terminal repeats (LTRs) comprising direct repeats of the U5, R, and U3 elements flank the internal protein-coding genes. A primer-binding site (PBS), which is necessary for reverse transcription can be found within the 5'-UTR region of the provirus.

1.1.1 ERVs

Endogenous retroviruses (ERVs) are a group of autonomous LTR retrotransposons in the human genome. They constitute around 7-8% of our genome and are believed to be the remnants of ancient germ-line retroviral infections (Belshaw et al., 2004; Jern and Coffin, 2008; Lander et al., 2001).

A typical ERV is made up of *gag*, *pol*, and *env* genes and is flanked by LTRs at both 3' and 5' ends (Figure 2). However, random mutations and deletions over the evolutionary timeline has led to either inactivation of the proviruses or restricted them to be intracellular pathogens (Ribet et al., 2008). As a result, their genomic size can vary from around several hundred base pairs to 12kb. Many of the ERV loci, however, still contain intact regulatory elements and can be transmitted vertically from one host generation to the next as a gene allele (Boeke and Stoye, 1997). Deletions of the *env* gene has rendered ERV an intracellular pathogen and has been proposed to have increased their retrotransposition potential making them 'superspreaders' (Magiorkinis et al., 2012) (Figure 3). While the working model seems plausible, there are also examples in which ancient retrotransposons have captured an *env* gene and successfully escaped from the host (Malik et al., 2000; Song et al., 1994). As a result, the exact origin of ERV is still being debated.

In an effort to understand the origin and age of ERVs in mammals, several methods have been deployed including looking at the integration histories and analysing host-virus co-phylogenetic relationships (Hayward, 2017). Integration analysis looks at the divergence between the 5' and 3' LTRs of a provirus in order to estimate the dates of ERV integrations. However, this method is restricted to recent integrants because sequence erosion of the older copies limits the usefulness of such



Adapted from Nature Reviews Microbiology, 10, Stoye, J. P., Studies of endogenous retroviruses reveal a continuing evolutionary saga, 395-406, Copyright 2012, with kind permission from Nature Publishing Group

Figure 3 Life cycle of ERVs.

Retroviral infection involves the following steps: binding to a specific receptor on the target cell surface; membrane fusion or receptor-mediated endocytosis; release of the viral core; reverse transcription; transit through the cytoplasm and nuclear entry; and integration into cellular DNA to give rise to a provirus. In the case of ERVs, deletion of the *env* gene means that the transcription (Pol II mediated) and translation of the viral capsid and enzyme proteins fuel the next round of reverse transcription and integration (known as retrotransposition) within the same host without the budding stage.

method. The alternative approach involves screening the host genome for orthologous ERVs. A shared ERV locus is strong indication of a common ancestor in which the integration event occurred. Using this approach, ERV-L elements have been found to be widespread among placental mammals and are estimated to have integrated into the genome ~104-110 million years ago (Lee et al., 2013a). The third method involves looking at the host-virus co-evolutionary relationships but this approach is restricted to retroviruses with relatively few cross-species transmissions such as spumaretroviruses (foamy viruses) (Switzer et al., 2005).

1.1.1.1 Life cycle of ERVs

Retroviruses such as human immunodeficiency virus (HIV) and murine leukaemia virus (MLV) are capable of reverse transcribing their positive-sense RNA genomes into DNA with the help of the reverse transcriptase enzyme (Hu and Hughes, 2012). The DNA is then integrated into the host genome with the help of a virally-encoded integrase enzyme (Craigie and Bushman, 2012). Once integrated into the germline, the provirus stays in the host and the sequence can be passed down vertically. The viral genes can then be transcribed and translated, with the help of host cellular machinery, producing proteins required for the assembly of new virus particles for as long as the infected cell is alive (Sundquist and Kräusslich, 2012). ERVs have a similar life cycle to their exogenous counterparts except for the extracellular steps potentially due to the deletion of *env* gene (Figure 3).

1.1.1.2 Classification

The lack of a clear classification and consensus nomenclature for ERVs has

complicated their study. Several methods have been used to classify ERVs. Historically, it was common to name a newly discovered ERV based on the RT sequence similarity to its exogenous counterpart. Under such a system, ERVs have been divided into three classes named Class I to III for Gammaretrovirus-like, Betaretrovirus-like and Spumaretrovirus-like elements, respectively (Jern et al., 2005). This method later proved to be over simplistic following the discovery of additional ERVs (Gifford et al., 2008; Katzourakis et al., 2007b). A second strategy has been to classify ERVs based on their primer binding site (PBS) sequences within their 5' untranslated regions (UTRs). The PBS sequence dictates the specific tRNA used by the virus for priming the reverse transcription of its RNA. Previously identified Betaretrovirus-like ERV groups, due to their similarity to MMTV, were then regrouped into the HERV-K super-group under the PBS classification method (Blomberg et al., 2009). This was later proved to be misleading with the availability of larger sequence datasets as not all elements under the HERV-K group have the same PBS sequence (Gifford and Tristem, 2003; Lavie et al., 2004). In addition to that, because ERV sequences are part of the genome, they can also be named as genetic loci. However, this has led to much confusion as they were named based on multiple criteria, for instance on their flanking sequence, chromosomal band or even the probe used for their cloning.

Sequence similarity between individual ERVs is another widely used method for grouping ERVs. In this case, the more conserved coding domain such as Pol is used as a template because it makes sequence recognition and alignment easier (Fitch, 1987). The same strategy can also be extended to include protein alignments using the same region. Alternatively, LTR sequences can also be useful to classify

ERVs including solo LTRs or copies devoid of Pol (Jern and Coffin, 2008; Jern et al., 2005). This is done by comparing a new LTR sequence to the database (such as Repbase) but it can get tricky when grouping recombined LTR sequences (Katzourakis et al., 2007a). Both methods essentially divide ERV sequences into LTR and internal sequences for analysis.

1.1.2 LINEs

LINEs are a group of autonomous non-LTR retrotransposons ubiquitously found in the human genome. While this class is made up of LINE1 (L1), LINE2 (L2) and LINE3 (L3) elements, L1 is the only group that is still active today (Lander et al., 2001). L1 is also the most successful TE in the human genome by mass and it constitutes approximately 17% of our genome resulting from its activities for the past 150 million years (Lander et al., 2001).

A canonical, full length human L1 is around 6kb in length. The element is comprised of an internal RNA polymerase II (Pol II) promoter at the 5'-untranslated region, two non-overlapping open reading frames (ORF1 and ORF2) separated by a short inter-ORF spacer and a 3' UTR containing a polyadenylation (poly-A) signal ending with an oligo dA-rich tail of variable length (Babushok and Kazazian, 2007) (Figure 4). ORF1 encodes a 40kDa RNA-binding protein (ORF1p) that shares little to no homology with known proteins, leaving its origins and its role in retrotransposition unknown. ORF2 encodes a 150kDa protein (ORF2p) with endonuclease (Feng et al., 1996) and reverse transcriptase (Mathias et al., 1991) activities (Babushok and Kazazian, 2007). The poly-A signal (AATAAA) within L1 is commonly read through during transcription resulting in 'chimeric' L1 transcripts

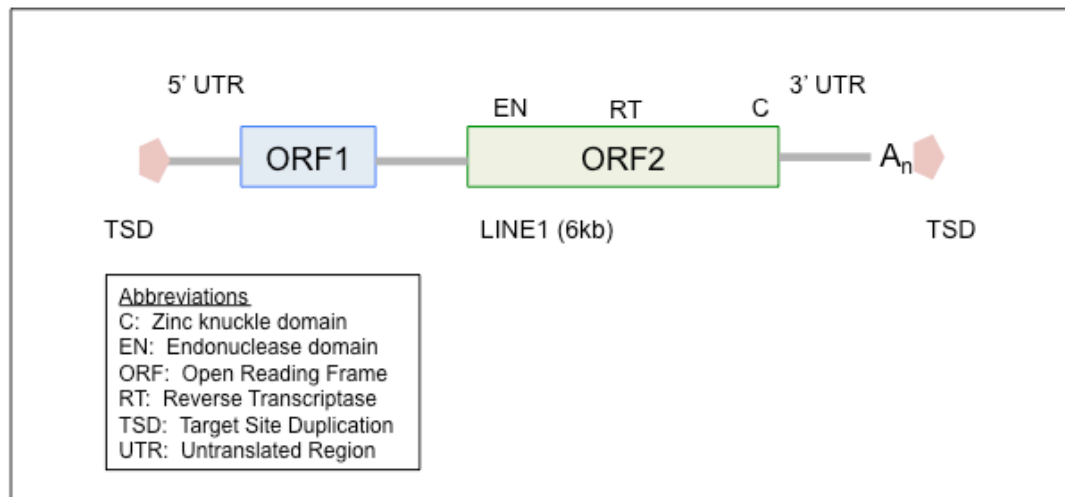


Figure 4 Structure of a LINE1 (L1) element.

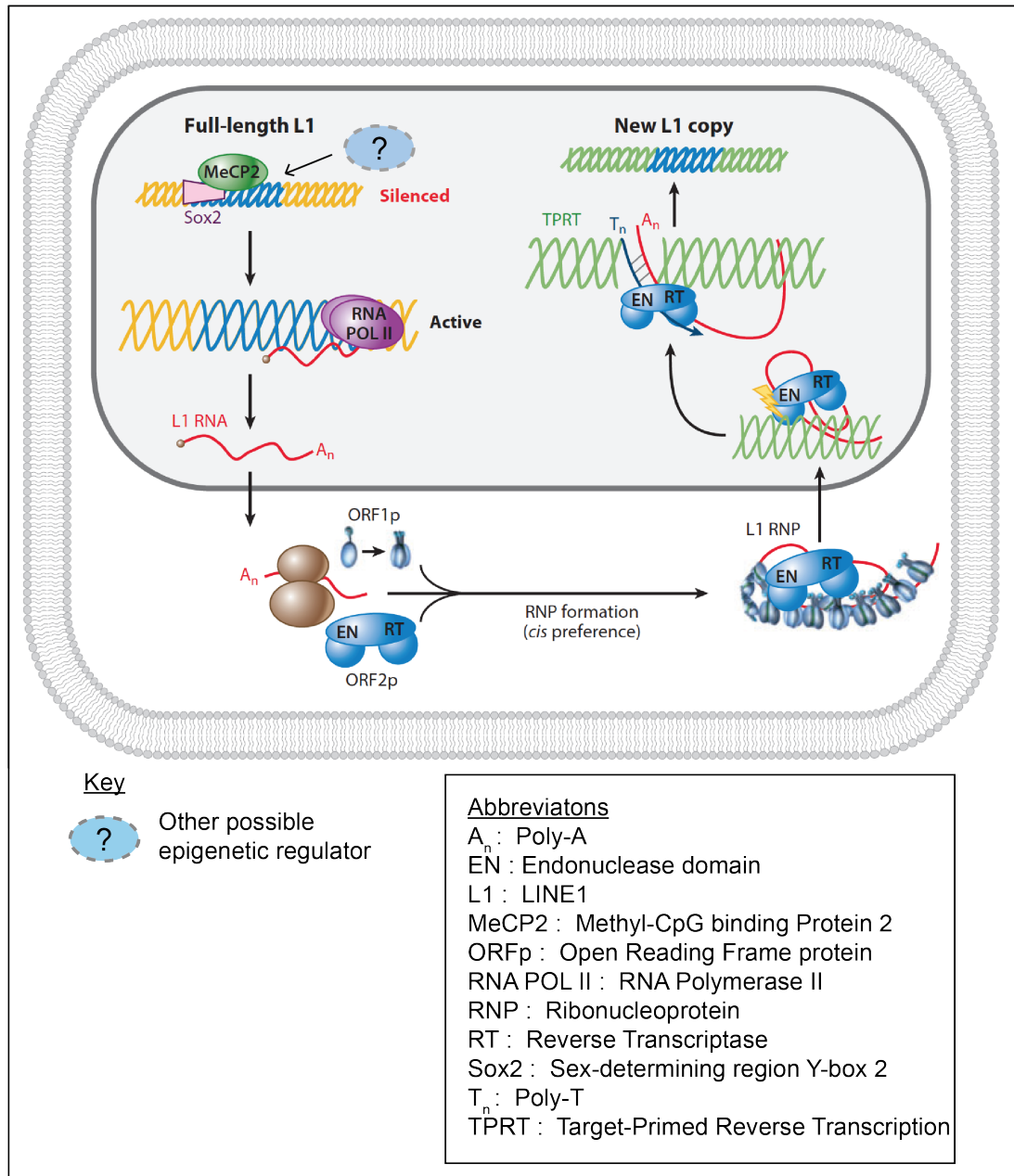
During integration, L1 elements induce target site duplications at both 5' and 3' ends. Abbreviations: A_n, poly-A.

composed of 3' flanking sequences and L1 sequences (Holmes et al., 1994; Moran et al., 1999; Moran et al., 1996).

1.1.2.1 Life cycle of LINE1

Out of the >500,000 L1 elements inserted in the human genome, less than 100 copies (<0.1%) are intact due to 5' truncations, internal rearrangements and mutations (Grimaldi et al., 1984; Lander et al., 2001; Ostertag and Kazazian, 2001; Szak et al., 2002). Most retrotransposition in the human population is believed to be the result of a number of highly active L1 loci (Brouha et al., 2003). The retrotransposition of L1 starts with the transcription at the internal Pol II promoter within the 5' UTR (Swergold, 1990). The L1's 5' UTR promoter is bidirectional and it can direct transcription in sense as well as antisense direction resulting in the production of chimeric transcripts containing both L1 5' UTR and genomic sequences upstream of the 5' end of that L1 locus (Nigumann et al., 2002; Speek, 2001). Translated ORF1p and ORF2p preferentially bind to the mRNAs from which they are derived, a phenomenon known as cis-preference (Kulpa and Moran, 2006; Wei et al., 2001) and this results in the formation of L1 ribonucleoprotein (RNP) particle (Hohjoh and Singer, 1996; Kulpa and Moran, 2005; Martin, 1991). The RNP particle then enters the nucleus (Kubo et al., 2006) where L1 target-site primed reverse transcription (TPRT) takes place (Cost et al., 2002; Feng et al., 1996; Luan et al., 1993).

During TPRT, ORF2p generates a single strand endonucleolytic break in the genomic DNA via its endonuclease activity to expose a 3'-OH (Feng et al., 1996). The liberated 3'-OH is then utilised as a primer by the reverse transcriptase to initiate



Adapted from the Annual Review of Genetics, 48, Richardson, S.R. et al., L1 retrotransposons and somatic mosaicism in the brain, 1-27, Copyright 2014, with kind permission from Annual Reviews

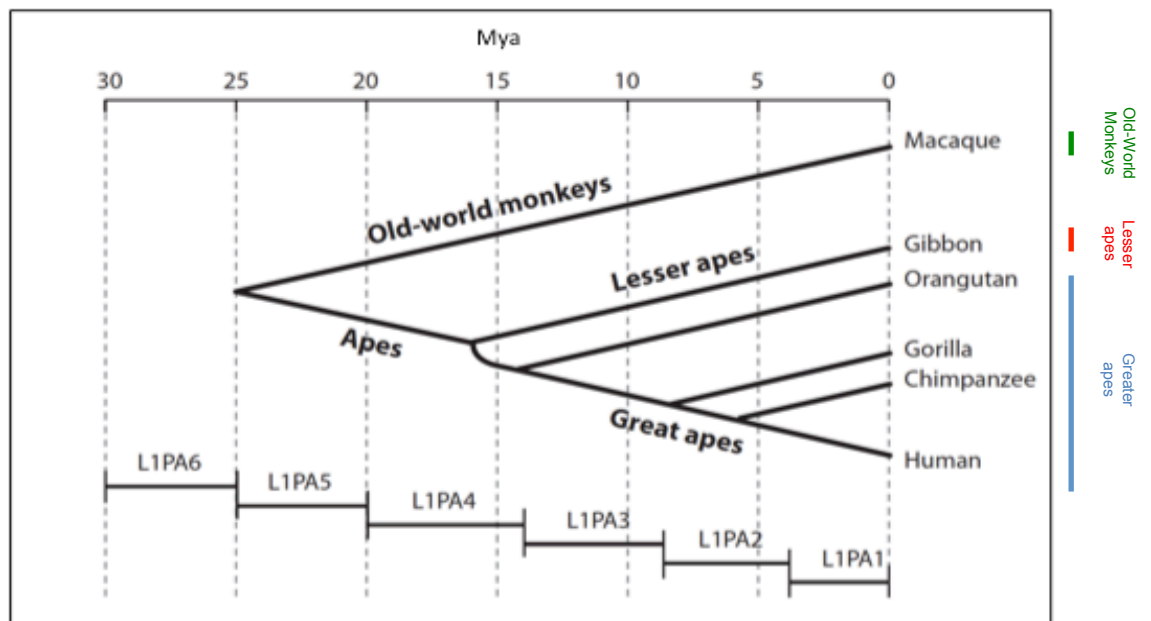
Figure 5 L1 retrotransposition mechanism.

Inactive L1 is either dead due to mutations or under the regulation of transcription factors in a tissue-specific manner. Factors such as MeCP2 and Sox2 have been demonstrated to regulate L1 in neuronal cell types. For L1 to retrotranspose, L1 RNA is first transcribed by RNA polymerase II and then translated. This will give rise to multiple copies of ORF1p and as few as one copy of ORF2p. Both proteins associate with their encoding RNA to assemble the L1 ribonucleoprotein particle. The RNP then enters the nucleus, where the L1 endonuclease activity nicks the genomic DNA at the consensus 5'-TTTT/A-3' to liberate a 3-hydroxyl (3'-OH) residue. The 3'-OH serves as the target site at which L1 reverse transcriptase initiates reverse transcription of its associated L1 mRNA. This process, known as target-site-primed reverse transcription, generates a new, frequently 5' truncated L1 insertion.

cDNA synthesis using the L1 mRNA as a template (Cost et al., 2002; Feng et al., 1996; Kulpa and Moran, 2006). The process generates a new L1 copy that is generally flanked by 7-20-bp target site duplications (TSDs) (Kazazian and Moran, 1998) (Figure 5). The molecular details behind TPRT have not been fully elucidated.

1.1.2.2 Classification

The replication mechanism adopted by L1 is not very efficient and results in defective copies with a truncation at their 5' end due to its inefficient reverse transcriptase but exact reason behind such strategy is unknown. In addition to this region, the first third of ORF1 and the 3' UTR evolve much faster than the other regions that presumably are crucial for L1's biochemical function. The rapidly evolving regions may either encode structures that are secondary for L1 replication or they need to be adapted for different host species (Furano et al., 2006). Therefore, the highly variable non-coding 3'-UTR, both in length and sequence, has been used in many of the diagnostic differences between L1 families within a species (Furano et al., 2006; Smit et al., 1995). The L1 elements also accumulate mutations at a neutral rate because the majority of L1 inserts are pseudogenes and as a consequence older families become more divergent than the younger ones (Boissinot et al., 2000; Hardies et al., 1986; Pascale et al., 1993; Voliva et al., 1984). This can be verified through the analysis on L1 families revealing that a single lineage of L1 families can amplify and evolve to replace its predecessor as the dominant family. This mode of evolution can be observed in human L1 families over the last 25 million years (Boissinot and Furano, 2001; Smit et al., 1995) (Figure 6).



Reprinted from Annual Review of Cell and Developmental Biology, 28, Thomas, C.A. *et al.*, LINE-1 retrotransposition in the nervous system, 555-573, copyright 2012, with kind permission from Annual Reviews

Figure 6 Evolution of L1.

Phylogenetic analyses of L1 in primates over the past 30 million years has shown the continuous demise of activity from the preceding L1 family that coincides with the arrival of a new L1 family. The nomenclature of L1 families depends on the species the family is present in, the global 3'-UTR sequence structure, a number to show the timing of the source gene within the group, and occasionally another lower case letter indicating that the family forms a side branch within the main lineage.

1.1.3 SINEs

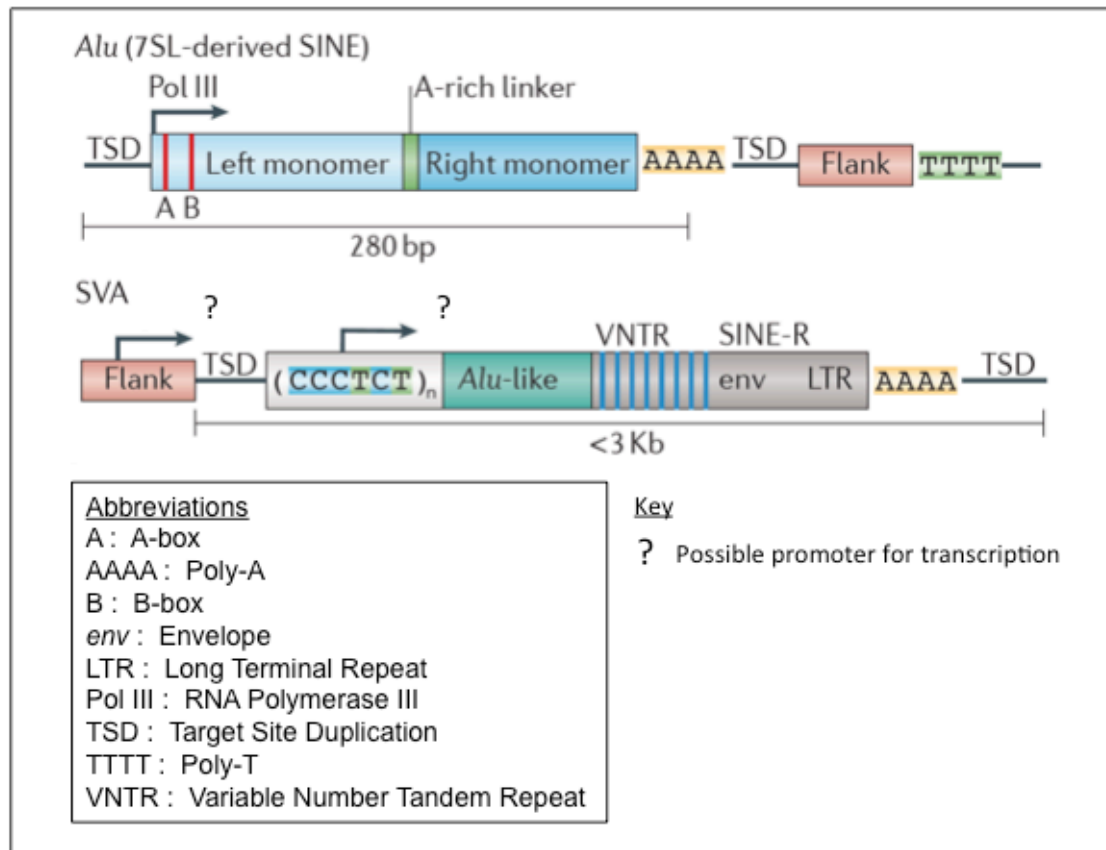
SINE is a collective term referring to short interspersed nuclear elements and they are non-autonomous and non-LTR retrotransposons. They are highly repetitive elements of typically 100-500 bps in length (Singer, 1982) and the largest SINE family within the human genome is *Alu*. There are over a million copies of *Alu* in the human genome making up >10% of the total genomic sequence and these elements are usually found in gene-rich regions (Consortium, 2005; Lander et al., 2001). *Alu* elements are typically ~300 bp in length and can be traced back at least ~65 million years (Batzner and Deininger, 2002).

1.1.3.1 Life cycle of SINEs

SINEs are frequently found within cellular genes and their defining feature is their transcription by Pol III from an internal promoter (Okada, 1991). In general, SINEs consist of two or more modules namely head, body and tail. The 5'-head originates from the cellular RNAs synthesised by Pol III. While most SINEs like mouse B2 has a head of tRNA origin (Daniels and Deininger, 1985), a few families such as human *Alu* have that of 7SL RNA (Ullu and Tschudi, 1984; Weiner, 1980), an RNA component of signal recognition particles or 5S ribosomal RNA (Kapitonov and Jurka, 2003). The origin of their body is usually unknown but some are derived from a partner LINE and mimic LINE RNA to hijack the reverse transcriptase for their cDNA synthesis (Okada et al., 1997). The SINE 3'-tail is a sequence of variable length consisting of simple repeats (Figure 7). To date, around 70 SINE families have been identified (Vassetzky and Kramerov, 2013) and their classifications are vastly dependent on the origin of their 'head' modules (Kramerov and Vassetzky, 2011).

SINEs do not encode proteins necessary for autonomous retrotransposition and are reliant on the reverse transcriptase and nuclease from LINEs (Moran et al., 1996). Their genetic constituents only contain an A-, a B- box sequence as well as a short sequence homologous to the 3' portion of their partner LINE within the same genome (Figure 3). The life cycle of SINEs starts with the transcription of its genome by Pol III. This is initiated by the recognition of the A- and B-boxes by a six-subunit protein complex known as TFIIIC. This protein-nucleotide interaction then leads to the binding of the TFIIIB complex, which then recruits the Pol III polymerase. Transcription starts from the upstream of A-box through the promoter region and continues to the site of the termination signal, which is a simple run of 4 or more thymidines (Ichiyanagi, 2013). In contrast, *Alu* elements possess no termination signals for Pol III, therefore, the transcription of *Alu* can extend downstream into the flanking sequence until a terminator is reached (Comeaux et al., 2009; Shaikh et al., 1997).

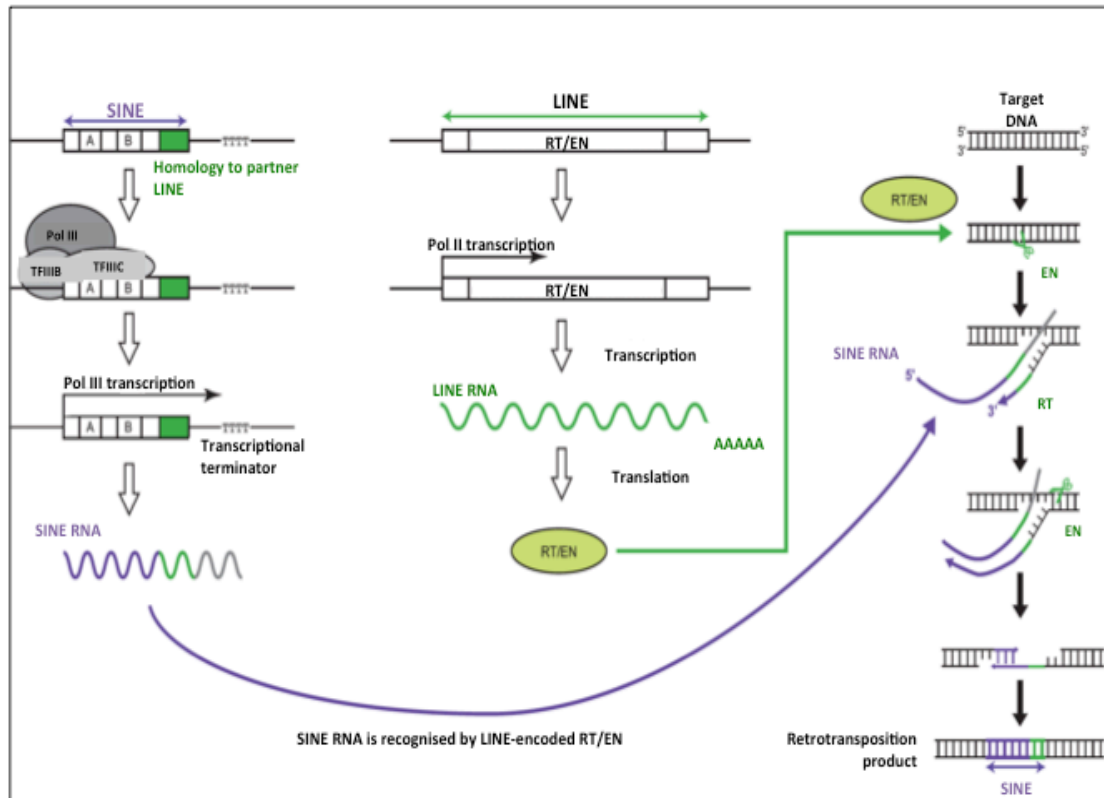
The homology between the 3' portion of the SINE structure and its LINE partner allows LINE reverse transcriptase to recognise the SINE RNA and initiate the target-primed reverse transcription (Luan et al., 1993) using the target genomic site cleaved by the LINE nuclease as a primer (Figure 8). For instance, human *Alu* RNA has a poly-A 3' sequence and this can be mobilised by the L1 reverse transcriptase, which recognises the poly-A region in the L1 RNA (Dewannieux et al., 2003). Despite the fact that L1's ORF1 and ORF2 proteins have a strong cis-preference for L1 RNA, *Alu* can hijack these proteins which are thus known as 'the parasite's parasite' (Weiner, 2002). The preservation of A- and B- boxes in SINEs allows them to retain their promoter sequence and capability of making another 'jump'.



Reprinted from Nature Review Neuroscience, 15, Erwin, J.A. et al., Mobile DNA elements in the generation of diversity and complexity in the brain, 497-506, Copyright 2014, with kind permission from Nature Publishing Group

Figure 7 Structure of *Alu* and SVA elements.

Both *Alu* and SVA elements hijack L1 machinery in order to retrotranspose. A full length *Alu* exhibits a dimeric structure formed by the fusion of two monomers originating from the signal recognition article 7SL RNA gene, sandwiching an A-rich linker region. A- and B- boxes harbour the internal promoter for RNA Pol III. *Alu* elements do not have a Pol III termination site, allowing for read-through transcription into the flanking sequence. Full-length SINE-VNTR-*Alu* (SVA) elements are a composite of multiple repeats, a CCCTCT repeat, two *Alu*-like sequences, a VNTR and a SINE-R region, which is homologous to the *env* and LTR sequences of an extinct HERV-K10. The lack of an internal promoter within the SVA elements means that they are reliant on flanking cellular promoters for their transcription. The 3' end of the element is a stretch of poly-A sequence.



Adapted from Genes & Genetic Systems, 88, Ichihayagi, K., Epigenetic regulation of transcription and possible functions of mammalian short interspersed elements, SINEs, 19-29, Copyright 2013, with kind permission from J-Stage

Abbreviations

A : A-box
 B : B-box
 EN : Endonuclease domain
 LINE : Long Interspersed Nuclear Element
 Pol II : RNA Polymerase II
 Pol III : RNA Polymerase III
 RT : Reverse Transcriptase
 SINE : Short Interspersed Nuclear Element
 TFIIB : Transcription Factor for polymerase III B
 TFIIC : Transcription Factor for polymerase III C

Figure 8 L1-mediated SINE retrotransposition mechanism.

To retrotranspose, SINE elements first produce a transcript through TFIIC binding to their A- and B-boxes which then recruits TFIIB and Pol III. Transcription stops at a downstream terminator, TTTT, occurring by chance. In parallel, a partner LINE encodes a protein with reverse transcriptase and endonuclease activities. The LINE protein cleaves the target DNA and initiates reverse transcription of SINE RNA in-trans. Recognition of the RNA template is mediated via the LINE homology region present in the 3' region of the SINE. SVA follows a similar pathway with the difference that its transcription depends on Pol II and requires an upstream cellular promoter.

1.1.3.2 SVAs

Another member of the non-autonomous retrotransposon group is the SINE-VNTR-Alu (SVA) elements. There are approximately 3,000 SVA elements resulting from their continued activities throughout the 25 million years of hominoid evolution (Ostertag et al., 2003; Wang et al., 2005). SVA elements are specific to the primate lineage (Cordaux and Batzer, 2009). Unlike a typical SINE, full-length SVAs are much larger in size and typically 2kb in length. SVA elements contain an hexamer repeat region, an Alu-like region, a variable number of tandem repeats region (VNTR), a HERV-K10-like region and a poly-A ending with an oligo dA-rich tail of variable length (Ostertag et al., 2003; Wang et al., 2005). These elements are also flanked by the L1-TSDs (Ono et al., 1987) (Figure 7).

In contrast to SINEs, SVAs are likely to be transcribed by Pol II and the resultant RNA is trans-mobilised by L1 retrotransposition machinery. In addition, SVA elements are likely to contain no internal promoter suggesting that they might rely on the promoter activity of their flanking gene promoter but this remains an open question (Beck et al., 2011; Ostertag et al., 2003; Wang et al., 2005) (Figure 8).

1.2 The impact of ERVs on their hosts

1.2.1 Exaptation of ERVs in host gene regulation

Despite their potential for insertional mutagenesis, no replicating ERV has thus far been identified in the human genome (Bock and Stoye, 2000; McPherson et al., 2001). Genome-wide studies have revealed ERVs to be a rich source of regulatory sequences that play a crucial role in the evolution of mammalian transcriptional networks

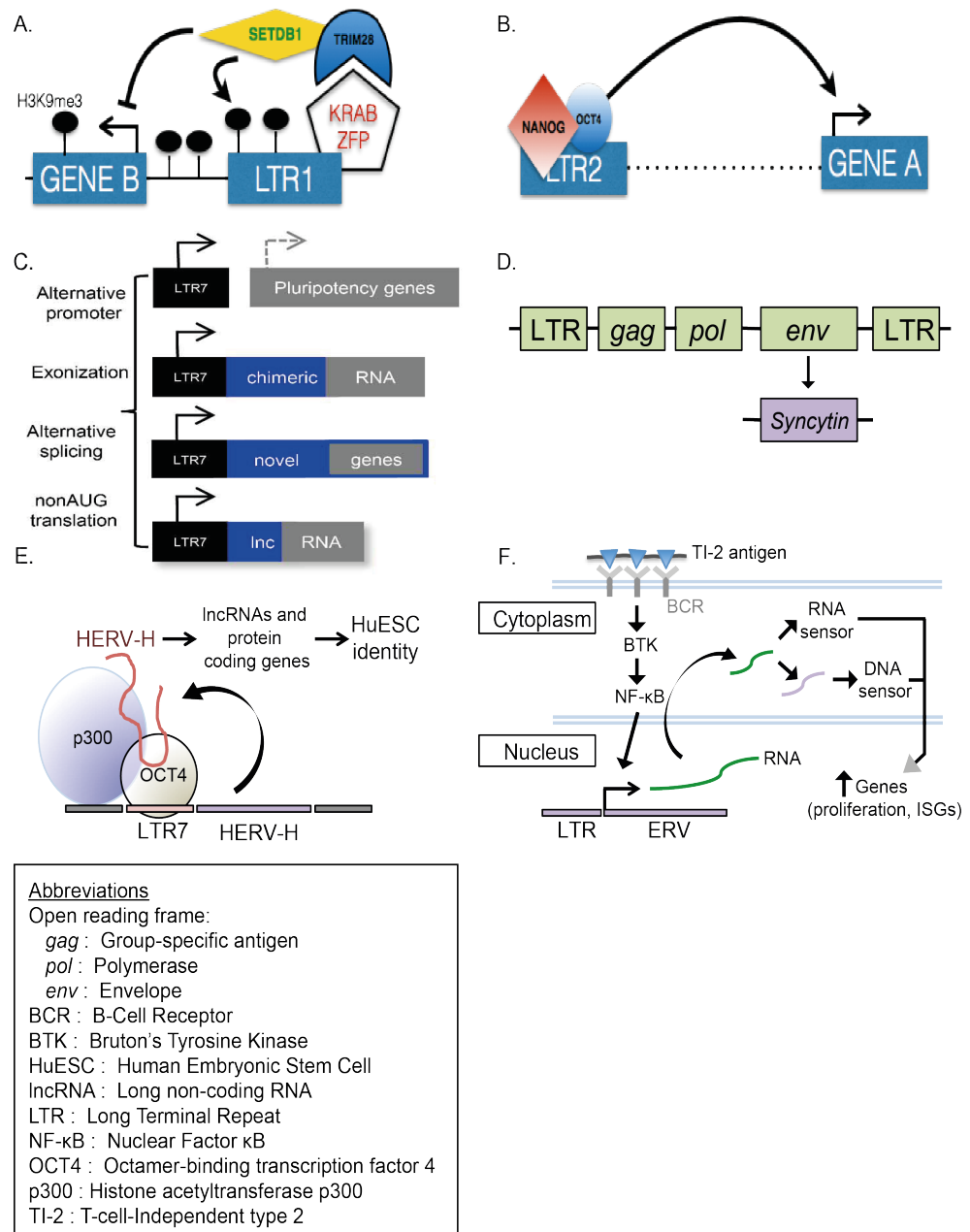


Figure 9 Mechanisms of ERV co-option.

A) Spreading of repressive histone marks at the site of a repressed ERV can impact on the expression level of cellular genes within the vicinity (Rowe et al., 2013a; Rowe et al., 2013b; Turelli et al., 2014). B) LTRs with binding sites for pluripotency transcription factors can serve as short-range or long-range enhancers to specific host genes (Suntsova et al., 2013). C) ERV elements expand the complexity of the ESC mRNA transcripts by functioning as alternative promoters to genes, being exonised in chimeric transcripts, providing alternative splice and acceptor sites, and driving non-AUG translation in order to give novel products (Wang et al., 2014). D) Host cells can exploit viral gene to devise novel gene product (Gifford et al., 2013; Mi et al., 2000). E) lncRNA transcribed from ERVs can interact with transcription factors in order to rewire the transcriptional network associated with the ground state of ES cells (Lu et al., 2014). F) Transcripts from ERVs can be sensed by the host as foreign and this can elicit a sensing event and a downstream signalling cascade that is crucial for normal immune function during T-independent B cell responses (Zeng et al., 2014).

(Brattas et al., 2017; Cordaux and Batzer, 2009; Feschotte and Gilbert, 2012; Gifford et al., 2013; Griffiths, 2001; Jern and Coffin, 2008; Katoh and Kurata, 2013; Manghera and Douville, 2013; Rebollo et al., 2012; Thompson et al., 2016) (Figure 9). The current population of ERVs is likely to reflect the loci that have been positively selected while those with a lethal phenotype have been out selected (Medstrand et al., 2002). There are also an estimated 577,000 solo LTRs in the human genome (Friedli and Trono, 2015), which arise from recombination between 5' and 3' LTRs of ERVs (Belshaw et al., 2007).

ERVs as alternative promoters

ERV LTRs have the intrinsic capacity to recruit cellular transcription factors and can be exapted as *cis*-regulatory elements with either activating or repressive effects (Chuong et al., 2013; Jacques et al., 2013; Sundaram et al., 2014). For instance, KZNF/ KAP1 and SETDB1 mediated repression of intact ERVs and solo LTRs can result in heterochromatin spreading and the repression of nearby genes (Karimi et al., 2011; Rebollo et al., 2011; Rowe et al., 2013b; Turelli et al., 2014; Wolf et al., 2015) (Figure 9). In contrast, some solo LTRs and incomplete ERVs can function as promoters or enhancers (Ecco et al., 2016; Friedli and Trono, 2015).

The promoter and enhancer functions of LTRs are often cell type specific. For example, in mice, LTRs of the MERV-L family (class III LTR retrotransposon) drive the expression of a cohort of stage-specific genes necessary for totipotency (Macfarlan et al., 2012; Peaston et al., 2004). In comparison, LTRs of MaLR and IAP families drive the expression of oocyte-specific transcripts in mice (Peaston et al., 2004; Veselovska et al., 2015). Similarly in humans, LTR7 from the primate-specific retrovirus HERV-H

has been demonstrated to drive many pluripotency-associated lncRNAs (Durruthy-Durruthy et al., 2016; Lu et al., 2014b; Wang et al., 2014b). Additionally, LTRs with promoter functions such as LTR3B, LTR14B, LTR12C, MLT2A1, THE1A and LTR5_Hs, are all expressed at discrete stages of human embryo development serving as promoters for distinct sets of gene (Goke et al., 2015).

Many LTRs are sequentially inactivated beyond zygote and two-cell stage (Hayashi et al., 2016; Hisada et al., 2012; Ishiuchi et al., 2015; Lu et al., 2014a; Macfarlan et al., 2012; Thompson et al., 2015) but some are positively selected to drive tissue-specific genes in differentiated cells (Figure 9). For example, MER39 and MER77 LTRs have been exapted to drive the *Prolactin* gene within endometrial cells in primates and rodents respectively (Emera et al., 2012; Romanish et al., 2007). The anti-apoptotic gene, *Naip*, has also been shown to be driven by MER12C and MT-C LTRs in rodents and primates respectively (Romanish et al., 2007). Additional examples of ERV-driven genes include the pancreatic amylase gene (Meisler and Ting, 1993) and germ cell-associated p63 (Beyer et al., 2011).

ERVs as lncRNAs

Genome-wide surveys have revealed that more than 75% of the ~10,000 annotated human lncRNAs contain TE sequences (Fort et al., 2014; Kannan et al., 2015; Kapusta et al., 2013; Kelley and Rinn, 2012) and they are enriched with LTR-elements when compared to other non-LTR elements (Kannan et al., 2015; Kapusta et al., 2013; Kelley and Rinn, 2012). While many LTR elements are transcribed as part of lncRNAs, some families act as promoters for lncRNAs. For instance, LTR7 drives the lncRNAs transcription in stem cell (Durruthy-Durruthy et al., 2016; Lu et al.,

2014b; Wang et al., 2014b) while HERV-H elements function in pluripotency-specific alternative splicing and alternative non-AUG usage (Goke et al., 2015; Wang et al., 2014b). Transcription from LTR7 is driven by the recruitment of pluripotency transcription factors including LTR binding protein-9 (LBP9), a transcriptional factor crucial for ground-state pluripotency, highlighting the role of LTR7 in maintaining the 'stemness' of stem cells (Ohnuki et al., 2014; Santoni et al., 2012; Wang et al., 2014b). Consistent with this role, increased expression of HERV-H lncRNA has been observed in induced pluripotent stem cells (iPSCs) (Ohnuki et al., 2014) and the loss of this RNA species compromises the ability of embryonic stem cells (ESCs) to self-renew (Lu et al., 2014b) (Figure 9). Analysis of human preimplantation embryos also revealed the expression of ERV1, ERV-K, and ERV-L derived noncoding RNAs but their functions remain elusive (Goke et al., 2015; Grow et al., 2015).

ERVs as enhancers

The enhancer function of ERVs is supported by the epigenetic enhancer signatures observed on these loci (Figure 9). Typical epigenomic signatures of an active enhancer include H3K4me1, H3K27ac, DNase I hypersensitivity, DNA hypomethylation, and the lack of the repressive H3K9me3 and H3K27me3 marks (Chuong et al., 2013; Fort et al., 2014; Jacques et al., 2013; Sundaram et al., 2014; Xie et al., 2013). One such example is the MaLR LTR, which acts as an enhancer for ~80% of mammalian *Pomc* gene expression (Lam et al., 2015). Similarly, LTR9 has also been shown to act as an enhancer for the *B-globin* gene cluster (Pi et al., 2010) and more recently, Chuong et al. have also shown that MER41 LTR enhances the expression of several interferon stimulated genes (Chuong et al., 2016).

The enhancer property of LTRs may be determined by using LTR reporter assays (Rowe et al., 2013b) and by their ability to drive enhancer RNAs (eRNAs) through bidirectional transcription (Kim et al., 2015; Plank and Dean, 2014). eRNA production by BGLII- and LTR17- derived enhancers has recently been shown to be present in pluripotent stem cells (Fort et al., 2014), and LTR9/LTR16A act as enhancers in certain tissues (Dunn et al., 2006).

ERVs as pathogen-associated molecular patterns (PAMPs)

Beyond their genomic sequences, ERV transcripts can also play a role in T-independent B cell responses (Zeng et al., 2014). When B cells are activated, ERVs get transcribed into mRNAs, which then get reverse transcribed into DNA. Both double stranded RNA and DNA are then sensed as PAMPs by retinoic acid-inducible gene I (RIG-I) and cyclic GMP-AMP synthase (cGAS) respectively (Zeng et al., 2014). These signaling events then result in B cell activation, proliferation and antibody production (Zeng et al., 2014). A similar mechanism (RNA sensing of ERVs) has recently been shown to take place in cancer cells treated with the DNA-demethylating agent, 5-Aza-2'-deoxycytidine (5-Aza) (Chiappinelli et al., 2015; Roulois et al., 2015).

1.2.2 ERVs and cellular protein expression

Novel cellular protein expression patterns

Studies in mice have shown that the repression of IAP elements, but not of L1 and the majority of the ETn/MusD elements, induces H3K9me3 and H4K20me3 repressive histone marks (Rebollo et al., 2011) and establishes a repressive chromatin environment in the vicinity (Elsässer et al., 2015; Sadic et al., 2015). The resulting spread of heterochromatin has been demonstrated to impact on the

expression of cellular genes such as *B3galt1* (Rebollo et al., 2011), which in turn will affect the protein expression profile of the cell. L1 has also been shown to play a role in the formation of facultative heterochromatin during X chromosome inactivation (Chow et al., 2016) and chromocenters (Kuznetsova et al., 2016) in mouse studies, although the exact mechanisms for these processes are not fully understood.

More recently, ERVs have been shown to impact on genes in neural progenitor cells (NPC) (Brattas et al., 2017; Fasching et al., 2015). These studies showed that, like in human and mouse ESCs (Rowe et al., 2013b; Turelli et al., 2014), the regulation of ERVs consequently leads to the establishment of local heterochromatin and the repression of neighbouring genes. Incidentally, some of these genes are involved in brain development (Brattas et al., 2017).

Beyond transcriptional regulation, ERVs have also been reported to benefit the host by providing new gene products and one such example is the syncytin gene, which plays a vital role in mammalian development and reproduction (Dupressoir et al., 2009; Mangeney et al., 2007; Mi et al., 2000). The syncytin family consists of Syncytin-1 and -2 and compelling evidence has been presented to show that they are derived from the envelope glycoprotein gene of the HERV-W family and are necessary for the formation of syncytiotrophoblasts in the placenta (Mi et al., 2000). Considering the importance of both properties to mammalian evolution, it is not surprising that *env* gene was captured and exapted independently on multiple occasions throughout mammalian evolution (Blaise et al., 2003; Cornelis et al., 2015).

There are several other examples of ERV protein products being expressed particularly during development. In the mouse genome, there are 16 copies of

MERV-L with intact open reading frames flanked by identical LTRs. MERV-L elements are active and their GAG proteins expressed from zygotic genome activation at the 2-cell stage up to the blastocyst stage (Kigami et al., 2003; Svoboda et al., 2004). By comparison, IAP transcripts and p73 GAG proteins are detectable first in the oocyte and expressed until their repression and DNA methylation (Pikó et al., 1984; Poznanski and Calarco, 1991; Svoboda et al., 2004). In contrast, MusD/ETn elements are transcribed in post-implantation embryos (Peaston et al., 2004). The function of ERV-derived proteins is unclear, except as antivirals (see below).

ERVs as antiviral agents

Budding of virus like particles (VLPs) has been observed in chicken and mouse embryos as well as in human blastocysts (Bieda et al., 2001; Dupressoir et al., 2005; Grow et al., 2015; Yotsuyanagi and Szollosi, 1981). Until recently, it was unclear if such VLPs have a physiological function in the cell but new data suggest they might play a role in protecting the embryos from certain infections. Grow et al. showed that the HERV-K accessory protein, Rec, could influence cellular mRNA translation and upregulate the classical virus restriction factor IFITM1 (Grow et al., 2015). Rec overexpression, however, has been suggested to also have a role in the formation of germ cell tumours by derepressing oncogenic transcription factors (Kaufmann et al., 2010).

The antiviral property of another viral proteins, Env, has been documented in multiple vertebrates. The earliest example of such property was observed in domestic chicken, in which endogenous avian leucosis virus (ALV) derived *env* genes were found to protect the animal from the insult of exogenous ALV by impeding its

entry (Payne and Chubb, 1968; Robinson et al., 1981). Similar observations were then reported in domestic sheep (endogenous Jaagsiekte sheep retrovirus (JSRV), enJSRV56A1) (Varela et al., 2009) and mice (Fv4 from endogenous MLV) (Gardner et al., 1991; Kozak, 2015). In humans, one candidate known as Suppressyn, an Env protein from HERV-F without the TM domain, has been proposed to regulate Syncytin-1 activity by interacting with the same cell receptor, ASCT2, thus blocking the binding of another domesticated *env* gene (Sugimoto et al., 2013). Several other retroviruses such as simian retroviruses 1 to 5, baboon endogenous virus, and feline RD114 all utilise the same receptor so the protein can potentially restrict these viruses in other species (Sommerfelt and Weiss, 1990). Another example in humans is the Env protein derived from a HERV-T provirus, which has recently been shown to be well preserved. A recent study proposed that ancient primates may have used this viral Env gene to restrict HERV-T entry causing extinction of the same element that provided the restriction factor (Blanco-Melo et al., 2017).

1.2.3 Negative effects of ERVs

Disruption of ERV regulation can have a negative impact on cellular gene expression and can ultimately cause disease. The association between ERVs and cancer has long been documented (Babaian and Mager, 2016; Kassiotis, 2014). In addition to the ERV accessory proteins, Np9 and Rec, most oncogenesis has been linked to the intrinsic regulatory capacity of the ERVs (Chen et al., 2013; Hohn et al., 2013; Kassiotis, 2014). For instance, inappropriate transcriptional activation of LTRs have been linked to multiple cancers in a phenomena known as ‘onco-exaptation’ (Babaian and Mager, 2016). Onco-exaptation occurs through aberrant promoter

activation or alternative promoter usage, which results in either ectopic or overexpression of an oncogene (Babaian and Mager, 2016).

To give an example, ERV activity has been linked to solid cancers such as breast cancer (Golan et al., 2008; Wang-Johanning et al., 2008) and melanoma (Buscher et al., 2005; Schmitt et al., 2013). In Hodgkin's Lymphoma (HL) patient samples, ectopic expression of the gene "*colony stimulating factor one receptor*" (CSF1R) was identified to be a possible cause of this cancer (Lamprecht et al., 2010). This protein is usually restricted to macrophages of the myeloid lineage but can get ectopically activated and expressed by a solitary THE1B LTR (MaLR-ERVL) in B cells (Lamprecht et al., 2010). The survival of HL cells is dependent on *Interferon Regulatory Factor 5 (IRF5)* gene expression driven by another primate LTR, LOR1a (Kreher et al., 2014). Furthermore, LTR elements of the ancient MaLR-ERVL class have also been shown to act as alternative promoters for the gene "*Erb-b2 receptor tyrosine kinase 4*" (ERBB4 or HER4) and give rise to truncated proteins with oncogenic properties (Scarfo et al., 2016).

LTR7-driven chimeric long non-coding RNAs are a characteristic of ESCs but their expression has been linked to colon cancer (Liang et al., 2012; Pérot et al., 2015). LTR7 and the neighbouring MER4C LTR drive the expression of the truncated form of *SLCO1B3* but its role in colon cancer is unclear for now (Liang et al., 2012; Pérot et al., 2015). Additionally, HERV-E-mediated ectopic expression of *fatty acid binding protein 7 (FABP7)* was found in several solid tumours such as aggressive breast cancer and some diffuse large B cell lymphoma (DLBCL) (Liu et al., 2012; Lock et al., 2014; Morin et al., 2011).

ERV-initiated non-coding RNAs have also been associated with cancers. One example of this is the association between *HOST2* lncRNA, which is produced from the LTR2B of a HERV-E element in epithelial ovarian cancer (Gao et al., 2015; Rangel et al., 2003). Also, the 5' end of the HERV-K element on chromosome 22q11.23 has been shown to fuse to the *ETV1* gene via chromosomal translocation and the resulting product is highly expressed in prostate cancer (Goering et al., 2015; Tomlins et al., 2007).

Beyond cancers, ERVs have also been linked with a variety of neurological diseases (Douville and Nath, 2014). One example would be amyotrophic lateral sclerosis (ALS) where the patient brain tissue presents elevated ERV-K activity (Douville et al., 2011). Animal models looking at motor dysfunction show increased ERV-K *env* gene expression in cortical and spinal neurons suggesting a potential cause of neurodegeneration (Li et al., 2015). ALS has also been linked to HERV-W *gag* and *env* expression, but whether this causes the disease remain debatable (Oluwole et al., 2007).

Multiple sclerosis (MS) has also been linked to ERV-W, ERV-W1, W2, and ERV-H/F elements (Christensen, 2016) following the observation of the presence of elevated Env antigen in the serum, and HERV-W DNA in peripheral blood mononuclear cells (PBMCs) from MS patients (Perron et al., 2012). Additionally, schizophrenia and other neuropsychiatric disorders have been linked to the hypermethylation of the HERV-W LTR located in the regulatory region of *GABBR1* (Hegyí, 2013), and ERV-K near the *PRODH* gene (Suntsova et al., 2013).

The link between autoimmune diseases and ERVs was first established following the discovery of conserved retroviral *pol*, with a homology to HERVs, in patient samples (Herrmann et al., 1998; Nelson et al., 1999). Since then, multiple mechanisms used by ERVs to induce autoimmune diseases have been proposed. Firstly, insertions near to immune genes can affect the expression level of genes and potentially contribute to the onset of certain diseases (Balada et al., 2009; Wu et al., 1993). Secondly, viral proteins from ERVs could be recognised by the host immune system as foreign, leading to antibody production. These antibodies might cross-react with certain self-proteins (Balada et al., 2009; Conrad et al., 1997) and could contribute to the pathogenesis of some autoimmune disease in a process known as molecular mimicry. Lastly, ERV proteins may also act as superantigens fuelling the expansion of autoreactive T-lymphocytes (Balada et al., 2009; Rolland et al., 2006).

The association between ERVs and systemic lupus erythematosus (SLE) was first implicated from SLE mouse models. GP70 (an endogenous retroviral Env protein) is expressed at high levels and anti-GP70 antibodies are detectable in lupus-prone mouse strains. However, this association was later dismissed with the discovery of non-autoimmune strains of mice with equally high GP70 level. Detection of autoantibodies in different populations of SLE patients has led to the proposal of multiple mechanisms explaining the causality of SLE including enhanced expression of *Fas* gene by early transposable element (ETn), molecular mimicry between HRES-1 Gag protein and host U1snRNP (Banki et al., 1992; Perl et al., 1995) and Rab4 proteins (Fernandez et al., 2009; Nagy et al., 2006), immune reactivity against HERV-K (Herve et al., 2002), HERV-H (Naito et al., 2003), and the ERV-3 Env protein (Li et al., 1996).

The association between insulin-dependent diabetes mellitus (IDDM) and HERV-K18 was first recorded following the detection of HERV-K antibodies (Herve et al., 2002) and HERV-K18 mRNA (Conrad et al., 1994) in the PBMCs and inflammatory lesions of patients. This observation can be explained by the detection of ENV superantigen encoded by HERVK-18 (Conrad et al., 1994; Marguerat et al., 2004) but ENV expression was later proved to be non-specific and similar observations could also be made in healthy participants (Murphy et al., 1998).

Patients with rheumatoid arthritis (RA) have been shown to express more HERV-K mRNA and antibodies (Herve et al., 2002) in their synovial fluid (Ehlhardt et al., 2006; Sicat et al., 2005). A different study has shown that higher levels of HERV-K10 *gag* mRNA can also be detected in both peripheral cells and synovial fluid cells of RA patients relative to the healthy controls (Ejtehadi et al., 2006). Interestingly, elevated transcript levels of the HERV-K18 superantigen were also reported in the synovial fluid mononuclear cells of juvenile RA patients (Sicat et al., 2005) but HERV-K (HML-2) *env* mRNA was downregulated in RA patients (Reynier et al., 2009).

The activity of ERVs in autoimmune diseases has also been linked to systemic sclerosis, autoimmune Addison's disease, alopecia areata, psoriasis (Perl, 2003), and even prion disease (Lee et al., 2013b). The question remains of whether ERVs are *bona fide* causative agents of these diseases or merely reflect the consequence of an inflammatory response in these diseases. To address this, more studies and better experimental models are needed in order to clarify the role of ERVs in various diseases.

1.3 Regulation of ERVs

Upon fertilization, the paternal genome undergoes active demethylation while the maternal genome is subjected to passive demethylation (Feng et al., 2010). This global loss of methylation is accompanied by a surge in transcriptional activity of certain ERV families (Feng et al., 2010; Oswald et al., 2000). Considering the potential impacts of ERV reactivation on host cells, it is not surprising that the host has evolved a spectrum of regulatory mechanisms. These strategies range from epigenetics, through to RNA interference and restriction factors.

1.3.1 Epigenetics

Chromatin regulation and DNA sequence both play a role in cell-type identity during development. Epigenetics refers to modifications on chromatin, rather than DNA sequence alterations, which lead to heritable effects on gene expression. Chromatin modifications include histone modifications and cytosine methylation and distinct epigenetic marks are associated with either an active or silent gene expression state. Chromatin signatures such as acetylation of histone 3 at the lysine residue at position 27 (H3K27ac) and monomethylation of histone 3 at lysine 4 (H3K4me1) are typically associated with active genes and enhancers (Rada-Iglesias et al., 2011); whereas trimethylation of histone 3 at lysine 9 (H3K9me3) and trimethylation of histone 3 at lysine 27 (H3K27me3) correlate with heterochromatin and gene silencing (Bulut-Karslioglu et al., 2014; Matsui et al., 2010; Walter et al., 2016). New ERV integrations can have two DNA-sequence driven effects on the host: Firstly, sequence integration itself alters the genetic content of our genome

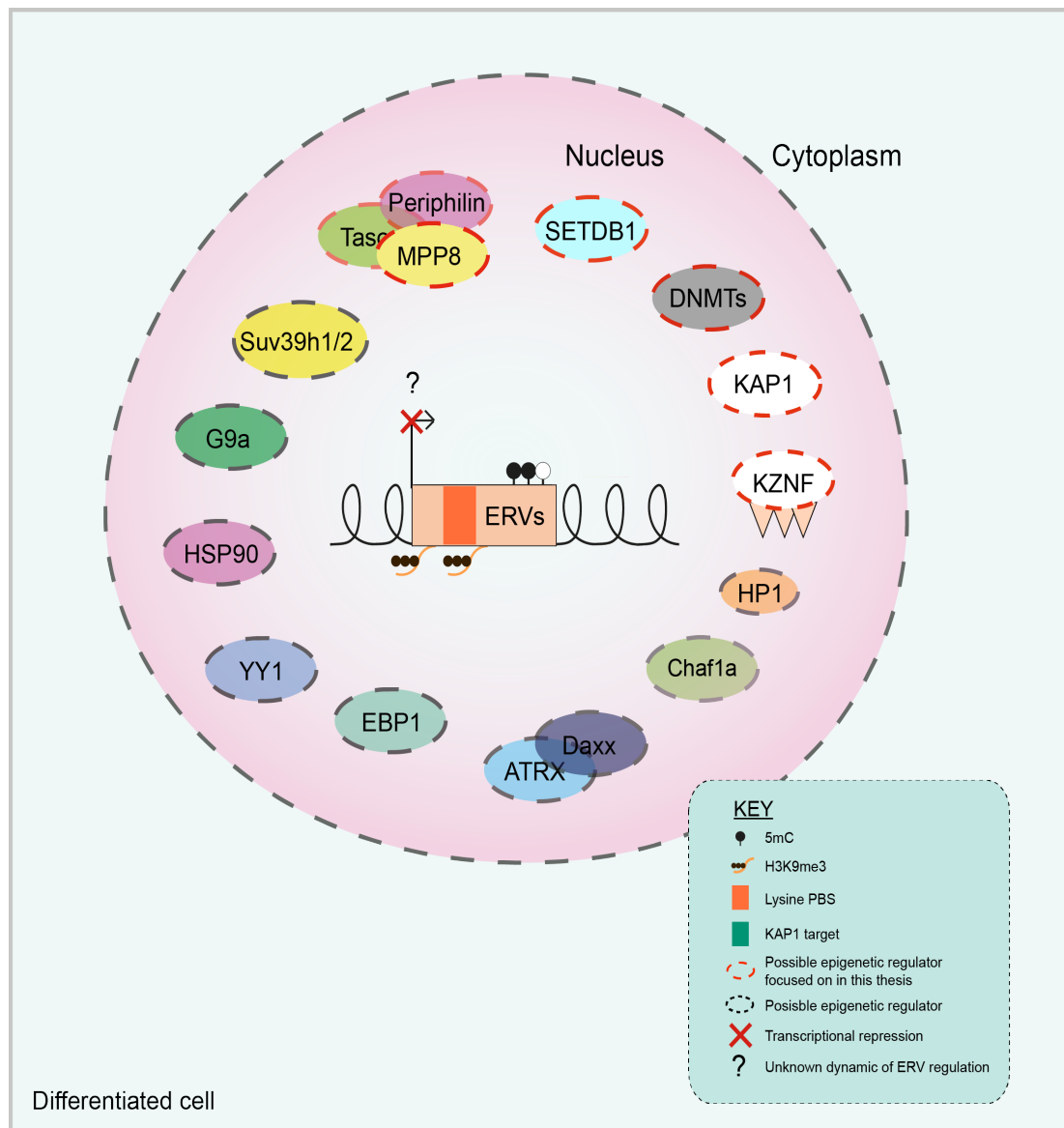


Figure 10 Putative epigenetic modifiers involved in ERV repression in differentiated cells.

KRAB-ZFPs (KZNFs) partner with KAP1, SETDB1 and DNMTs to silence ERVs in ESCs and early development but it is unclear which of these factors are required to maintain ERV repression in differentiated human cells. It is also unknown whether the recently described HUSH complex is necessary for ERV repression.

and secondly, the transcription factor binding sites within the ERV regulatory elements can orchestrate epigenetic traits heritable in the daughter cells.

1.3.2 Epigenetic regulation of ERVs in early development

KZNFs and KAP1

Transcriptional regulation of ERVs includes DNA methylation and histone methylation. However, little is known about how these silent chromatin marks are targeted to ERVs early in development. One known mechanism utilizes Kruppel-associated-box zinc finger proteins (KZNFs) and KAP1 (Rowe et al., 2010; Turelli et al., 2014) (Figure 10). This is supported by data from KAP1 and KZNF knockouts in early development, which have demonstrated their role in repressing retrotransposons (Brattas et al., 2017; Ecco et al., 2016; Fasching et al., 2015; Rowe et al., 2010; Wolf et al., 2015). One of these studies also showed that the KAP1 and KZNF pathway plays a minor role in retrotransposon repression in mouse liver following the liver-specific knockout of KAP1 (Ecco et al., 2016).

The first line of evidence implicating KAP1 in ERV repression was based on murine leukemia virus (MLV) repression in murine embryonic stem cells (Wolf and Goff, 2007; Wolf and Goff, 2009). Indeed, Daniel Wolf showed that MLV restriction is initiated by recognition of the 18 nucleotide-long primer binding site Pro (PBS-Pro) sequence by Zfp809 followed by KAP1 recruitment. This recognition is highly specific such that a single mutation is sufficient to disrupt the interaction and thus MLV repression. Although KZNF/KAP1 repression is highly conserved between mice and humans, ERV populations are often species-specific. This observation is in line with the finding that KZNF repertoires are species-specific and many have undergone positive selection in parallel with the rise of new classes of ERVs in the genome

(Corsinotti et al., 2013; Emerson and Thomas, 2009; Lukic et al., 2014; Najafabadi et al., 2015; Thomas and Schneider, 2011).

To initiate ERV repression, KZNFs must first bind to their DNA targets through their arrays of C2H2-type zinc finger domains. The C-terminal regions of KRAB-ZNFs contain tandemly arranged arrays of C2H2 zinc finger modules, comprising a few to more than 30 fingers. Each of these fingers is separated from each other by a highly conserved linker sequence and can interact with three nucleotides (Looman et al., 2002). Their conserved N-terminal KRAB domains recruit KAP1 protein to the recognized sequence and give specificity to the repression pathway (Najafabadi et al., 2015).

The human genome encodes around 350 KZNFs, the majority of which are found in clusters, a result of their evolution through duplication and deletion of zinc finger modules (Vogel et al., 2006). The family is fast evolving and has recently been traced back as far as the common ancestor of coelacanths, lungfish and tetrapods (Imbeault et al., 2017). The unique repertoire in different species, however, suggests that they have evolved to perform species-specific transcriptional regulation and / or that different ones perform the same functions across species (Huntley et al., 2006). Of note, around half of the human KZNFs are widely conserved and their functions are mostly unknown, although presumably they bind to conserved targets. Large arrays of zinc finger proteins (ZNFs) can be viewed as a panel of effector proteins with specific DNA targets analogous to an antibody repertoire specific for foreign antigens. Many of the binding targets for human KZNFs have been identified through chromatin immunoprecipitation experiments (Imbeault et al., 2017; Najafabadi et al., 2015; Schmitges et al., 2016) but only a few, namely ZNF91 and ZNF93, have

been functional characterized, in this case to repress SVA and L1 subfamily sequences (Jacobs et al., 2014). Apart from SETDB1, the identities of the other enzymes required for KAP1 repression of SVA and LINE1 remain incompletely characterized, although a recent study in mouse ESCs has revealed other potential candidates such as Chaf1a (Yang et al., 2015).

KAP1 is a ubiquitously expressed protein implicated in numerous processes including cell differentiation, pluripotency maintenance, tumor development, cell cycle and apoptosis (Iyengar and Farnham, 2011). It is composed of a RING-B-box-Coiled-coil (RBCC) domain at its N-terminal, a heterochromatin protein 1 (HP1) binding domain, a plant homeodomain (PHD) and a bromodomain at its C-terminal (Cammass et al., 2000). The trimerisation of the RBCC domain encapsulates the KRAB domain in a protease-resistant core to promote the folding of the KRAB domain, which enables an interaction between the two domains (Peng et al., 2000). The other domains are involved in the recruitment of downstream silencing complexes including SETDB1, heterochromatin proteins of the HP1 family and the NuRD histone deacetylase complex (Ayyanathan et al., 2003; Cammass et al., 2004; Ivanov et al., 2007; Peng et al., 2000; Satou et al., 2001; Sripathy et al., 2006; Underhill et al., 2000; Zeng et al., 2008). The central HP1 binding domain contains a hydrophobic PxVxL pentapeptide and its interaction with the chromodomain domain of HP1 proteins has been shown to be required for gene silencing (Nielsen et al., 1999). The rest of the central region is highly extended and flexible, possibly providing KAP1 the adaptability to interact with a plethora of protein complexes (Lechner et al., 2000). The bromodomain of KAP1 can interact with the backbone of histone tails and, together with the PHD domain, interact with the NuRD complex and SETDB1 (Schultz

et al., 2002; Schultz et al., 2001). Its interaction with SETDB1 is dependent on sumoylation at lysines 554, 575, 676, 750, 779, and 804 of the bromodomain (Zeng et al., 2008) and this is mediated by the intramolecular E3 ligase activity of the PHD domain (Ivanov et al., 2007). The sumoylation of KAP1 residues stimulates the histone methyltransferase activity of SETDB1 and the resulting H3K9me3 then creates a high affinity binding site for the KAP1-HP1 complex (Ivanov et al., 2007). These interactions provide a mechanistic basis for the deposition of repressive chromatin marks and the spread of heterochromatin (Fietze et al., 2010).

YY1 and EBP1

Zinc finger protein Yin Yang 1 (YY1) and ErbB3-binding protein (EBP1) have been implicated in the MLV silencing complex (Schlesinger et al., 2013; Wang et al., 2014a) (Figure 10). To mediate repression, YY1 needs to bind to the LTR of both exogenous and endogenous retroviruses in ESCs while phosphorylation of tyrosine residue 383 of the protein can abrogate repression by interrupting YY1 binding to DNA and RNA (Wang and Goff, 2015). Intriguingly, loss of YY1 can reactivate retroviral genomes and leads to the removal of H3K9me3 but not H3K27me3 histone marks from the loci (Schlesinger et al., 2013). YY1-mediated restriction and its interaction with KAP1 is restricted to ESCs and this pathway plays little or no role in differentiated cells (Schlesinger et al., 2013). EBP1 was previously demonstrated to be part of the ZFP809-KAP1 repression complex but its exact role in MLV silencing and ERV regulation requires further study (Wang et al., 2014a).

Heat Shock Protein 90

Heat Shock Protein (HSP90) was shown to utilise the KAP1-mediated epigenetic pathway to insulate genes from the regulatory influence of nearby ERV loci (Hummel et al., 2017; Rowe et al., 2013b) (Figure 10). Its effect on KAP1-regulated ERVs can be observed outside of ESCs in differentiated macrophages but its mechanism of action is unknown (Hummel et al., 2017).

Histone modifications

Histone N-terminal tails can be targeted for chemical modifications including acetylation, phosphorylation, methylation and ubiquitination. These modifications directly impact on chromatin structure and on the access to DNA binding proteins, which in turn affects gene expression. Histone modifications play an important role in gene regulation as well as in retrotransposon control. Histone modifications mediated by proteins such as histone deacetylases (HDACs), histone methyltransferases (ESET/SETDB1, G9a and Suv420H1/2), lysine-specific histone demethylase 1A (KDM1a/ LSD1), polycomb complexes and the arginine methyltransferase PRMT5 have all been implicated in retrotransposon repression (Garcia-Perez et al., 2010; Kim et al., 2014; Leeb et al., 2010; Leung et al., 2014; Leung et al., 2011; Macfarlan et al., 2011; Maksakova et al., 2013; Matsui et al., 2010; Reichmann et al., 2012; Rowe et al., 2013a; Rowe et al., 2013b; Yang et al., 2015). Of note, the role of histone methylation in retrotransposon repression is often linked to DNA methylation so it is crucial to understand the interplay between these two types of modifications.

H3K9me3 is the most common repressive histone mark observed at LTR

elements and can be deposited by all histone methyltransferases (HMTases) that target H3K9 (Martens et al., 2005; Mikkelsen et al., 2007). For instance, Suv39h1 and Suv39h2 are involved in trimethylation of H3K9 at satellite repeats and knockout of these proteins is lethal for embryos at the E12.5 stage (Lehnertz et al., 2003; Peters et al., 2001) (Figure 10). Another HMTase, G9a, is involved in the deposition of H3K9me1 and H3K9me2 marks. While the loss of G9a does not impact on repressive H3K9me3 marks at retrotransposons, DNA methylation is significantly reduced following G9a knockout, indicating a separate role in DNA methylation maintenance. This has been shown to be due to its recruitment of DNMTs (Dong et al., 2008; Tachibana et al., 2008) (Figure 10). In addition, the loss of G9a leads to only a mild upregulation of IAP elements suggesting that histone methylation is sufficient to silence certain retrotransposons in the absence of DNA methylation. SETDB1 has directly been shown to play a role in repressing ERVs in both ESCs and somatic tissues (Collins et al., 2015; Jahner et al., 1982; Matsui et al., 2010) (Figure 11). In support of this being a main function of SETDB1, its knockout in mouse ESCs results in elevated expression of MusD and IAP elements (Matsui et al., 2010). Repressed retrotransposons are often enriched for multiple histone marks such as H4K20me3, H3K9me3 and H3K27me3 (Kourmouli et al., 2004; Martens et al., 2005; Mikkelsen et al., 2007) suggesting a coordination between different epigenetic modifiers in ERV repression. For instance, the overlap between H3K9me3 and H3K27me3 suggests a potential role of Polycomb group proteins as well as Setdb1 in ERV silencing (Golding et al., 2010; Leeb et al., 2010).

The co-factor HP1 is made of two structured domains, a chromodomain and a chromoshadow domain, connected by an unstructured hinge. The chromodomain

interacts with KAP1 and H3K9me3 and is therefore necessary for ERV repression (Hiragami-Hamada et al., 2016; Lachner et al., 2001; Maksakova et al., 2013; Sripathy et al., 2006) and the dimerisation of the protein is involved in chromatin compaction (Bannister et al., 2001; Lachner et al., 2001; Smothers and Henikoff, 2000). Upon binding to methylated nucleosomes, HP1 undergoes a switch from an auto-inhibitory state to a spreading competent state (Canzio et al., 2013) and this interaction has also been observed in differentiated cells (Sharma et al., 2012; Wolf et al., 2008a).

H3.3 and HUSH

Recently, the chaperone complex containing histone variant H3.3, deposited by α -thalassaemia/mental retardation syndrome X-linked (ATRX) and death-domain-associated protein (DAXX), was shown to be required for ERV repression in mouse ESCs (Elsässer et al., 2015; Sadic et al., 2015) (Figure 10). This study shows that H3.3 is enriched at class I and class II ERVs (particularly ETN/MusD and IAP elements) and deposition at these elements is dependent on ATRX and DAXX (Elsässer et al., 2015). However, whether H3.3 plays any functional role in ERV repression is controversial (Wolf et al., 2017).

The Human silencing hub (HUSH) is another recently described complex implicated in the silencing of newly integrated DNA in the host genome and may potentially play a role in ERV repression (Tchasovnikarova et al., 2015) (Figure 10). This complex is composed of Tasor, MPP8, and periphilin proteins and it is conserved from fish to humans (Tchasovnikarova et al., 2015). The HUSH complex is recruited to genomic loci by interacting with H3K9me3 resulting in more Setdb1 recruitment, which in turn increases the levels of H3K9me3 marks in a positive feedback loop (Li

et al., 2011; Tchasovnikarova et al., 2017; Tchasovnikarova et al., 2015). An independent study has reported an interaction between the chromodomain of MPP8, dimethylated DNMT3A (at lysine residue 44) (Dnmt3ak44me2) by G9a or GLP protein (Chang et al., 2011) leading to the formation of Dnmt3a-MPP8-G9a/GLP complex, providing a possible link between histone modifications and DNA methylation (Chang et al., 2011).

Histone deacetylation

Histone deacetylation occurs on histone 3 and histone 4 tails at lysine residues and plays a role in the repression of IAP elements. Consistent with this, inhibiting histone deacetylases (HDAC) has been connected to the reactivation of virus-like 30S elements in mice (Chen and Townes, 2000; Lorincz et al., 2000; Schübeler et al., 2000). In addition, chemical modifications can also act cooperatively with other epigenetic modifications such as histone methylation. For instance, components of the CoRest complex, which includes LSD1 (or KDM1A), HDAC1 and HDAC2, can mediate the removal of H3K4me1 and H3K4me2 and loss of histone acetylation cooperatively to create a repressive chromatin environment (Shi et al., 2004; You et al., 2001). DNA methylation has also been linked to histone deacetylation through an interaction between methyl-CpG-binding protein (MeCP2) and deacetylase complexes (Jones et al., 1998; Nan et al., 1998).

DNA methylation

DNA methylation of retrotransposons is re-established following global demethylation and zygotic genome activation in cleavage embryos. DNA methylation is mediated by DNA methyltransferases (DNMTs), which comprise the four

members: DNMT3a, DNMT3b, DNMT1 and DNMT3L (Figure 10). DNMT3a and DNMT3b mediate *de novo* DNA methylation, while DNMT1 protein maintains DNA methylation in dividing cells and DNMT3L, which lacks enzymatic activity, is important for targeting the *de novo* methylation complex involving DNMT3a and DNMT3b (Bestor et al., 1988; Bestor, 1988; Chedin et al., 2002; Ooi et al., 2007; Suetake et al., 2004). In mice, the loss of DNMT3b has been shown to result in increased activity of IAP and endogenous MLV elements but the effect is more dramatic upon loss of DNMT1 (Okano et al., 1998). A DNMT3a-DNMT3L complex has been reported to bind to unmethylated H3K4 tails, thereby targeting DNA methylation away from active regions that are marked by mono-methylated H3K4 (Ooi et al., 2007). After *de novo* methylation, ERV silencing is dependent on the maintenance of methylation at these loci and the loss of DNMT1 or DNA methylation can lead to embryonic lethality and aberrant immune activation in cancer cells (Chiappinelli et al., 2015; Roulois et al., 2015; Walsh et al., 1998). The mechanism of DNA methylation maintenance by DNMT1 involves binding to hemi-methylated DNA and H3K9me2/3 in a UHRF1 (also known as NP95 or ICBP90) - dependent manner (Liu et al., 2013). In pre-implantation embryos, however, DNMT1 is involved in protecting certain IAP1 insertions and imprinted regions from global DNA demethylation but it is unclear how DNMT1 is preferentially maintained at these regions (Lane et al., 2003; Reik, 2007).

Lymphoid-specific helicase 1 (LSH1) has also been reported to be important for the methylation of elements such as IAPs, L1s and SINE B1s in mice (Dennis et al., 2001). LSH1 knockout leads to the loss of repressive histone marks and derepression of retrotransposons without affecting neighbouring genes, indicating the specificity

of LSH1-mediated repression (Huang et al., 2004; Yan et al., 2003). This pathway has previously been linked to DNMT3a and DNMT3b-mediated methylation by an unknown mechanism (Zhu et al., 2006).

1.3.3 ERV regulation in somatic cells

It is known that embryonic stem cells possess intrinsic factors that gradually repress retroviruses and part of this silencing activity is linked to the KZNF / KAP1 pathway (Friedli et al., 2014; Rowe et al., 2010; Teich et al., 1977; Wolf and Goff, 2007; Wolf and Goff, 2009; Wolf et al., 2008b). This has been documented in both mouse and human ES cells and is consistent with *in vivo* data (Grow et al., 2015; Macfarlan et al., 2012; Rowe et al., 2010). Beyond the blastocyst stage, ERV sequences are thought to regain DNA methylation, allowing their potential permanent inactivation in somatic tissues (Yoder et al., 1997) in which KAP1 may be redundant (Martens et al., 2005; Meissner et al., 2008; Mikkelsen et al., 2007; Rowe et al., 2013a; Rowe et al., 2010). DNA methylation becomes the dominant silencing mechanism later in development, for example in E9.5 embryos (Walsh et al., 1998) and in murine embryonic fibroblasts that are isolated at E13 (Rowe et al., 2013a).

Longstanding parallel data, however, has documented widespread transcription of ERVs in multiple tissues including in blood and brain (Krieg et al., 1992; Muotri et al., 2005; Reilly et al., 2013) strongly suggesting that expression of certain ERVs is tissue-specific similar to some cellular genes (Ecco et al., 2016; Fort et al., 2014; Goke et al., 2015). Indeed, KAP1 has been reported to regulate some ERVs and related genes in neural progenitor cells and even in differentiated cells including mouse liver (Brattas et al., 2017; Ecco et al., 2016; Fasching et al., 2015).

Setdb1 has been shown to repress ERVs in mouse B-lymphocytes in collaboration with KAP1 (Collins et al., 2015) suggesting KAP1 plays a continuous role in differentiated tissues.

1.3.4 Small RNA-mediated repression

In plants, small RNA-mediated gene silencing via *de novo* methylation is well established (Law and Jacobsen, 2010) and the mechanism appears relevant for ERV repression because it involves sequence-specificity. Small RNA-mediated gene regulation includes Piwi-interacting RNA (piRNA), short interfering RNA (siRNA), and microRNA (miRNA).

piRNAs are 23-30 nucleotides long and they are derived from repeat-rich clusters and this process is independent on dicer. This pathway is well-described in *Drosophila* in which antisense piRNAs bind to Piwi or Aubergine proteins to cleave sense mRNAs from retrotransposons (Aravin et al., 2007). Piwi-like proteins known as Miwi, Mili and Miwi2 are found in the germ line (Aravin et al., 2007; Girard et al., 2006). The resulting sense piRNAs can bind Ago3 proteins, which then cleave antisense mRNAs. This “ping-pong” mechanism amplifies the piRNAs and cleaves the retrotransposon mRNAs resulting in silencing. In mice, the same pathway is involved in the cleavage of retrotransposon mRNAs and loss of these proteins has been linked to the activation of IAP and L1 elements.

Unlike piRNAs, the biogenesis of siRNA is Dicer-dependent and this pathway is well characterized in *Drosophila* (Ghildiyal and Zamore, 2009). siRNA production is initiated by the recognition of intracellular, long, double-stranded RNAs (dsRNAs) by Dicer-2, which then cleaves them into siRNAs (Weber et al., 2006). siRNAs bind Ago2

but with only one of the duplex and the resulting siRNA-Ago2 RNA-induced silencing complex (RISC) recognises complementary transcripts and cleaves them (Bronkhorst and van Rij, 2014; Ghildiyal and Zamore, 2009). A proportion of siRNAs are of endogenous origin (endo-siRNAs) and can therefore be directed towards retrotransposons (Ghildiyal and Zamore, 2009). There is also supporting evidence showing that siRNAs play a role in the repression of both IAPs and L1 in mouse ES cells (Svoboda et al., 2004; Yang and Kazazian, 2006).

A third class of small RNAs is known as miRNAs that are synthesized in a Dicer-dependent manner (Ghildiyal and Zamore, 2009). miRNAs are encoded in the genome and derived from primary miRNA transcripts (pri-miRNAs), which are then cleaved into shorter precursors (pre-miRNAs) by Drosha protein (Ghildiyal and Zamore, 2009; Siomi et al., 2011). Pre-miRNAs are then processed by Dicer-1 and the mechanism of degrading complementary transcripts is mediated by Ago1 (Ghildiyal and Zamore, 2009).

Despite reports of small RNAs repressing retrotransposons or exogenous pathogens, the precise importance of their role in repression in human cells remains relatively unexplored.

1.3.5 Restriction factors

A network of restriction factors including APOBEC3s, SAMHD1, Fv1 and Trex1 have evolved to protect the host cell from exogenous retroviral (XRVs) assaults (Best et al., 1996; Lahouassa et al., 2012; Sheehy et al., 2002; Yan et al., 2010). The similarity of life cycle between ERVs and XRVs means that these factors can also restrict ERVs, which is presumably their primary purpose.

APOBEC3 proteins work by interfering with the processivity of reverse transcriptase and induce hypermutation of viral cDNAs by cytidine deamination. APOBEC3 has been shown to restrict the activities of L1, Alu, IAP1 and MusD elements (Esnault et al., 2005). APOBEC3s can theoretically act at the post-transcriptional stage of the ERV life-cycle and can introduce mutations to retrotransposons rendering them inactive.

Trex1 is a 3'-5' DNA exonuclease that plays a role in preventing the accumulation of reverse-transcribed cDNA in the cytoplasm (Yan et al., 2010). This mechanism has been demonstrated for both HIV-1 and endogenous retroelements (Stetson et al., 2008; Yan et al., 2010), the latter which was shown using TREX1 KO mice. When TREX1 was overexpressed, it reduced the retrotransposition of L1 and IAP elements *in vitro* while *Trex1* knockout (KO) led to an accumulation of single stranded DNA (ssDNA) from L1, SINEs and ERVs within heart cells of KO mice (Stetson et al., 2008).

The role of SAMHD1 as a restriction factor was first reported in the context of HIV-1 infection via a mechanism involving the control of dNTP pools in immune cells (Ryoo et al., 2014). This is through SAMHD1's ability to catalyse deoxynucleotides (dNTPs) into deoxyribonucleosides and triphosphates, rendering them unavailable for DNA replication and reverse transcription (Ryoo et al., 2014). SAMHD1 has been shown to inhibit L1 retrotransposition through stress granule formation but its importance in controlling the activity of ERVs remains unknown (Zhao et al., 2013).

Restriction factors can also be derived from endogenous Env or Gag proteins produced from ERV elements themselves, which is the case for the Fv1 protein. Fv1 resembles a MERV-L Gag and it can block MLV infection at a post-reverse

transcription stage (Best et al., 1996). This block is likely to be mediated through capsid interaction and such protection offered by ERVs is similar to the recently proposed role of HERV-K Rec protein (Grow et al., 2015; Kaiser et al., 2007).

1.4 ERVs and the immune system

The life cycle of ERVs involves proviral transcription, export of the mRNA and translation of viral proteins. Transcripts can then be reverse transcribed into cDNA in the cytosol resulting in their potential integration into the host genome. Interestingly, nascent viral particles have been observed budding from the plasma membrane although these have been shown to be non-infectious retroviral-like particles (RLPs) (Bieda et al., 2001; Grow et al., 2015). Despite ERVs only very rarely giving rise to infectious viruses, ERVs have been linked to immune modulation and autoimmune diseases.

Given the exaptation of ERVs over millions of years, it has been proposed that the host should have developed sufficient immune tolerance to ERV proteins and see ERVs as 'self' (Tugnet et al., 2013). However, the connection between autoimmune diseases and ERVs challenges this idea (Section 1.2.3). In addition, ERV-induced innate immune activation in cancer cells as well as in immune cells suggests that these elements still retain their 'foreign' characteristics (Chiappinelli et al., 2015; Goel et al., 2017; Roulois et al., 2015; Zeng et al., 2014). One mechanism by which ERVs can trigger an immune response is by producing nucleic acids (both RNA and DNA) that resemble pathogen-associated molecular patterns (PAMPs) thereby activating pattern recognition receptors (PRRs). PRR activation invokes a series of complex signalling pathways, which results in the production of pro-inflammatory

cytokines, chemokines and type I interferons (IFN- α and IFN- β). The causative contribution of ERVs to different disease phenotypes is unclear and how the host sees certain ERVs as self and others as foreign also requires further investigation. ERV proteins might cause damage or cellular stress to host cells (Tang et al., 2012), which in turn gives rise to danger-associated molecular patterns (DAMPs) that would activate similar classes of PRRs and IFN signalling (Tang et al., 2012).

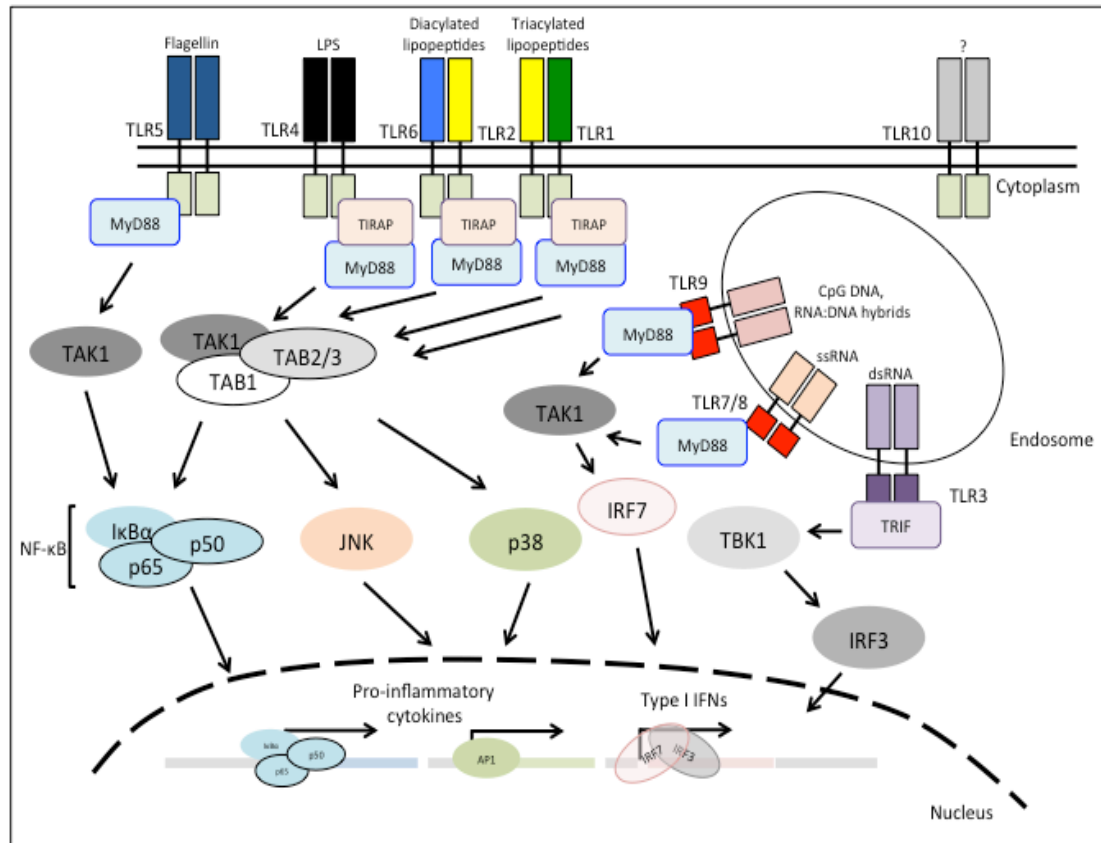
1.4.1 Toll-like receptors

Toll-like receptors (TLRs) are transmembrane proteins found either on the plasma membrane or on endosomal membranes and there are 10 human TLRs, each with different subcellular localisations and cognate ligands (Lester and Li, 2014) (Figure 12). TLRs are made up of an extracellular ligand-binding domain with leucine-rich repeats and a conserved cytosolic domain comprising a Toll-IL-1 resistance (TIR) domain (O'Neill et al., 2013). Upon ligand binding, signalling from the TLRs involves dimerization and the recruitment of TIR domain-containing adaptor molecules, such as myeloid differentiation factor 88 (MyD88) to the TIR domains (Lester and Li, 2014). This then activates a cascade of intracellular kinases and ubiquitin ligases resulting in the nuclear translocation of transcription factors such as NF- κ B and IRF 3 and 7 (O'Neill et al., 2013) (Figure 11), which in turn activates the expression of proinflammatory cytokines and type I IFNs (Figure 11).

TLRs 1, 2, 4, and 5 are all localised to the plasma membrane. TLR2 can recognise bacterial lipopeptides by heterodimerisation with TLR1 to sense triacylated lipopeptides and with TLR6 to detect diacylated lipopeptides (O'Neill et al., 2013) (Figure 11). TLR2 has also been shown to recognise a wide variety of non-lipopeptide

PAMPs (Kawai and Akira, 2009). TLR4 acts as a sensor for the bacterial product lipopolysaccharide (LPS) while TLR5 has been shown to sense bacterial flagellin (O'Neill et al., 2013) (Figure 11). On the other hand, TLRs 3, 7, 8, and 9 are localised to the endosomal lumen and bind nucleic acid ligands (Figure 11). TLR3 senses dsRNA, TLR7 and 8 detect ssRNA, and TLR9 recognises unmethylated CpG DNA and RNA:DNA hybrids (Gürtler and Bowie, 2013; Rigby et al., 2014) (Figure 11). The ligand for TLR10 has not yet been identified but the sequence of the receptor is known to be most similar to that of TLR1 and may have the capability to heterodimerise with TLR2 (O'Neill et al., 2013). TLR2 has been implicated in the induction of NF- κ B in psoriasis and this is linked to single-nucleotide polymorphisms in HERV-K dUTPase (Ariza and Williams, 2011). This HERV-K mediated immune activation also results in a Th1 and Th17 cytokine response (Ariza and Williams, 2011). Despite this link, whether the UTPase is detected by TLR2 directly remains unknown because TLR2 is expressed on the cell surface while HERV-K proteins are expressed intracellularly.

TLR4 and CD14 have been reported to interact with the Env protein of HERV-W in the context of multiple sclerosis (Rolland et al., 2005; Rolland et al., 2006). This interaction results in the production of pro-inflammatory cytokines including IL-1 β , IL-6 and TNF- α (Rolland et al., 2006) in addition to activating dendritic cells and inducing a Th1 response, potentially linking innate and adaptive aspects of the disease.



Adapted from the Journal of General Virology, 96, Hurst, T.P. and Magiorkinis, G., Activation of the innate immune response by endogenous retroviruses, 1207-1218, Copyright 2015, with kind permission from the Microbiology Society

Figure 11 TLR signalling in response to ligand binding.

TLRs on the cell surface (TLR1, 2, 4, 5 and 6) typically detect foreign lipid or protein ligands while endosomal TLRs (TLR3, 7, 8 and 9) sense nucleic acids. The signalling pathway has been simplified to show all the members of the TLR family as well as their respective ligands. The signalling pathways for TLRs involve multiple kinases in order to link the initial sensing to a resulting pro-inflammatory state. Following ligand binding, TLR4 recruits MyD88 using its TIR domain. This then leads to the recruitment of IRAKs and activation of TRAF6, an E3 ubiquitin ligase for TAK1. Ubiquitinated TAK1 then forms a complex with TAB1 and TAB2. This activates TAK1, which then phosphorylates IκB kinase and the MAPK kinases and sequentially activates NF-κB followed by MAPKs. MAPK kinase then activates JNK and p38. Activated JNK and p38 activate AP-1 to drive the expression of pro-inflammatory cytokines. Activated NF-κB also participates in the production of cytokines during inflammation. The ligand for TLR10 is currently unknown.

TLR7 and TLR9 are both linked to the production of the ERV envelope glycoprotein and gp70, an autoantigen postulated to function in the pathogenesis of SLE (Yoshinobu et al., 2009). In a lupus-prone mouse model, TLR7 and 9 have been demonstrated to induce the production of serum gp70 and contribute to inflammation in the mouse model (Baudino et al., 2010). Another line of evidence comes from the deficiency in IRF5 a specific signalling pathway used by TLR7 and TLR9 in the mouse model. This deficiency leads to an increase in gp70 *gag* and *env* mRNA suggesting that IRF5, TLR7, and TLR9 are involved in the control of murine ERV expression (Yu et al., 2012). Furthermore, TLR7 has also been shown to facilitate antibody production and class switching in mice (Browne, 2011) and a TLR7 deficiency results in the failure of antibody production that is required to control ERV viraemia (Yu et al., 2012).

1.4.2 RNA sensing

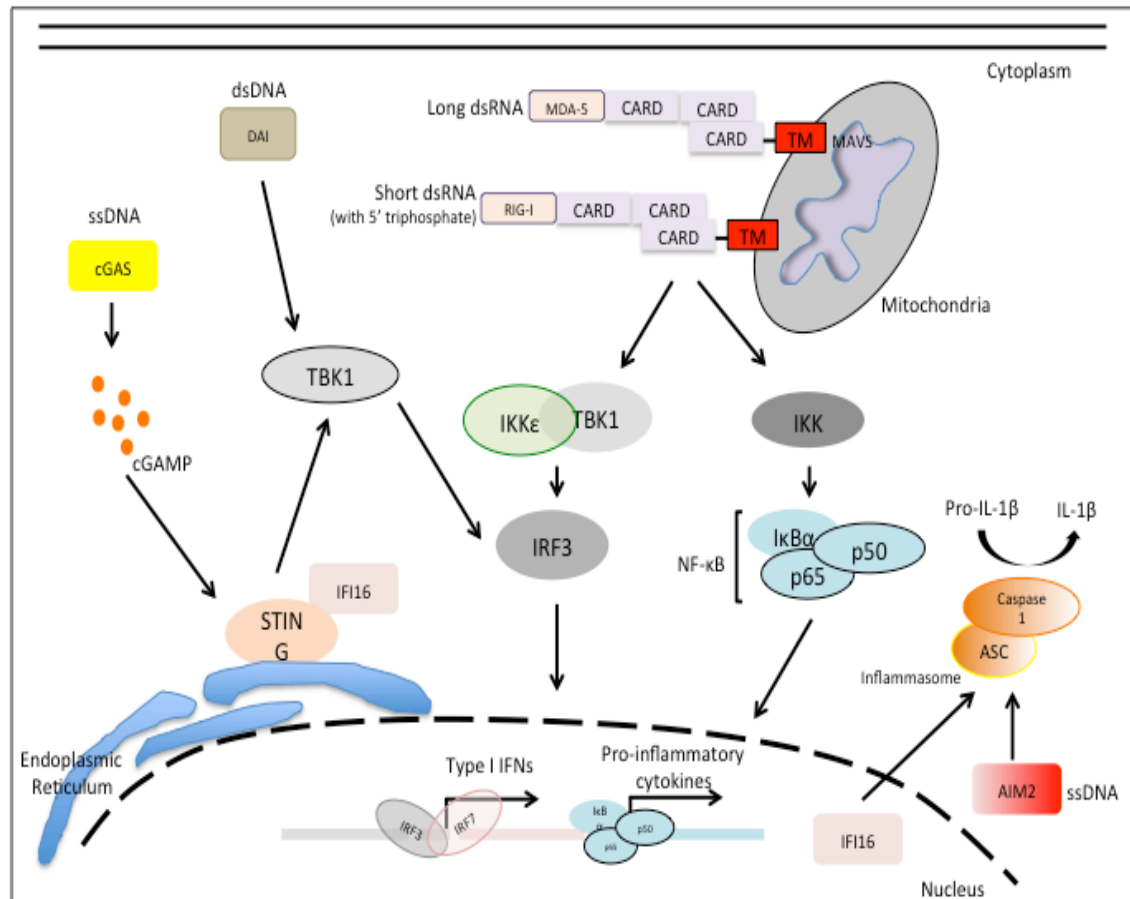
In addition to the TLRs, cytosolic RNAs are sensed by a series of PRRs including retinoic acid-inducible gene I (RIG-I) and melanoma differentiation-associated gene 5 (MDA5). These belong to the RIG-I-like receptor (RLR) family, a class of RNA helicases that recognise viral RNA in the cytoplasm and induce the production of type I IFNs (Gürtler and Bowie, 2013). RIG-I senses short dsRNA and ssRNA with a 5'-triphosphate moiety while MDA5 detects long dsRNA molecules (Kato et al., 2008) (Figure 12). Following detection, RIG-I and MDA5 signal through their N-terminal homotypic CARD interaction with the mitochondrial outer-membrane-localised adaptor molecule mitochondrial antiviral signalling (MAVS) (Gürtler and Bowie, 2013) (Figure 12). This leads to the phosphorylation of

transcription factors IRF 3 and 7. The phosphorylated IRFs then translocate to the nucleus and bind to the IFN-stimulated regulatory element (ISRE) and induce IFN expression along with other antiviral factors (Hiscott, 2007) (Figure 12).

Evidence for cytosolic ERV RNA sensing has been lacking but two independent studies have recently shown that ERV RNA can activate innate immune signalling in cancer cells (Chiappinelli et al., 2015; Goel et al., 2017; Roulois et al., 2015). These studies utilised the DNA methyltransferase inhibitor, 5-Aza to reveal that 5-Aza-driven immunotherapy likely depends on ERVs. The studies showed that the anti-tumour efficacy of the drugs is mediated through viral mimicry through dsRNA derived partly from ERVs. Immune activation depended on TLR3 and MDA5/MAVS recognition pathways as well as IRF7, leading to Type III IFN production and apoptosis (Chiappinelli et al., 2015; Goel et al., 2017; Roulois et al., 2015). Sensing of ERV RNA by MAVS has also been previously documented to play a role in T-cell independent B-cell responses in mice (Zeng et al., 2014). These studies point to the immunostimulatory nature of ERV transcripts, which thus need to be metabolised in order to prevent unwanted innate immune activation. The mechanism by which the host metabolises cytosolic dsRNA is also relevant, therefore, in preventing ERV-mediated immune activation. For instance, the 3'-5' RNA SKIV2L helicase found in exosomes has been shown to degrade immunostimulatory endogenous RNA molecules although its role in metabolising ERV transcripts remains unknown (Eckard et al., 2014).

1.4.3 DNA sensing

Cells have also evolved a collection of sensors for cytosolic DNA PAMPs. The absent in melanoma 2 (AIM2)-like receptor has been demonstrated to sense dsDNA (Hornung et al., 2009), while IFN- γ -inducible protein 16 (IFI16) can detect ssDNA (Unterholzner et al., 2010) (Figure 12). ssDNA detected by IFI16 has been reported to have a stable stem loop structure (Jakobsen et al., 2013) (Figure 12). Following cytosolic DNA detection, IFI16 recruits the adaptor stimulator of IFN genes (STING), resulting in IFN- β expression (Unterholzner et al., 2010) (Figure 12). More recent studies, however, have proposed that it plays a major role as an adaptor to the DNA sensing pathway that proceeds through cGAS activation (Almine et al., 2017). Alternatively, DNA-dependent activator of IFN regulatory factors (DAI) has also been demonstrated to detect long dsDNA (Takaoka et al., 2007). Importantly, cGAS has been demonstrated to detect DNA from retroviruses (Gao et al., 2013). DNA sensing by cGAS results in the production of 2'3' cyclic-AMP-GMP (cGAMP), a secondary messenger molecule, which then activates STING and IRF3 for IFN induction (Schoggins et al., 2014; Sun et al., 2013) (Figure 12). cGAS is the best characterised DNA sensor to date and is plays a major role in cytosolic DNA sensing (Sun et al., 2013). The role of ERV DNA in stimulating the innate immune response first came from a study in mice looking at the 3' repair exonuclease 1 (TREX1) (Paludan and Bowie, 2013). TREX1 is upregulated by the expression of type I IFN and NF- κ B and it plays a role in regulating endogenous DNA PAMPs and DAMPs (Xu et al., 2014). TREX1 knockout mice induce an IRF3-dependent immune response to cytosolic DNA, which is thought to include ERV cDNA, in a cGAS- and STING- dependent manner (Ablasser et al., 2014). In addition, the requirement of cGAS to sense ERV cDNA was



Adapted from the Journal of General Virology, 96, Hurst, T.P. and Maglorkinis, G., Activation of the innate immune response by endogenous retroviruses, 1207-1218, Copyright 2015, with kind permission from the Microbiology Society

Figure 12 Cytosolic pathogen-recognition receptor (PRR) signalling.

Cytosolic RNA sensors interact with MAVS through CARD-CARD domain interactions. MAVS is anchored to the outer membrane of mitochondria and this localisation is crucial for the transduction of pro-inflammatory signals during the activation of NF-κB and IRF3. DNA sensors such as cGAS and IFI16 signal through STING, which is localised to the endoplasmic reticulum. IFI16 can also be found in the nucleus and its translocation to the cytosol is involved in the formation of the inflammasome. Production of the Inflammasome activates caspase 1-mediated cleavage of pro-IL-1β. The formation of the inflammasome can also be induced by AIM2 following ssDNA detection.

also demonstrated in T-cell independent B-cell activation (Zeng et al., 2014). These observations suggest that, at least in mice, some ERVs have the capability to get reverse transcribed into cDNA and stimulate host innate immunity. Despite this, the identity of the ERVs responsible for these sensing events and whether they encode functional reverse transcriptases is currently unknown.

1.4.4 Consequences of innate immune activation

During infection, innate immune activation results in the production of pro-inflammatory mediators including IFNs, cytokines and chemokines and establishes an antiviral state to counteract infection until it is cleared. However, in the event of retrotransposon-induced immune activation, the presence of ERV transcripts may continue to stimulate the immune system leading to chronic inflammation potentially contributing to autoimmunity. Chronic induction of potent IFNs can have pathological effects on the cells. For instance, IFNs can mediate autocrine and paracrine signalling leading to an antiviral state, in which affected cells have low levels of cellular metabolic processes and ultimately go into apoptosis (Zuniga et al., 2007). Improper clearance of apoptotic cells can present another source of autoantigens and exacerbate inflammation (Szondy et al., 2014). Type I IFNs stimulate natural killer cells and clonally expand cytotoxic T cells, linking innate with adaptive immunity (Zuniga et al., 2007). In a lupus model, self-nucleic acid sensing by TLRs correlates with the production of autoantibodies providing an example of how ERV detection by PRRs can potentiate the production of autoantibodies (Barrat et al., 2005).

During inflammation, IFN secretion can create a positive-feedback loop whereby the transcription factors that get activated during a change of disease state can bind to and activate the transcription of more ERV loci. *In silico* studies of the HERV-K 5'LTR has led to identification of multiple binding sites for pro-inflammatory transcription factors such as NF- κ B (Manghera and Douville, 2013). NF- κ B is commonly activated during an inflammatory response its binding to the HERV LTR can potentially upregulate HERV transcription. Indeed, this has been demonstrated using LPS and TNF- α treatments, in which NF- κ B induction results in the increased expression of HERV-K, HERV-H, and HERV-W mRNAs (Johnston et al., 2001). Such interaction between ERVs and inflammatory transcription factors provides a mechanism through which endogenous elements can sustain inflammation and contribute to disease progression.

In contrast, several ERV viral products have been demonstrated to have an immunosuppressive effect on host immunity including immunosurveillance evasion of cancer cells (Kassiotis, 2014). For example, HERV-FRD Env protein can suppress the maternal immune response to fetal alloantigens while HERV-H Env protein has been found to play a role in preventing immune rejection of tumours in mice (Mangeney et al., 2001; Mangeney et al., 2007). In addition, HERV-K Env protein has been found to contain an immunosuppressive domain and can induce the production of IL-10, an immunosuppressive cytokine (Morozov et al., 2013).

1.4.5 ERVs and adaptive immunity

As discussed, it is unclear if ERVs are sensed as “self” or ‘foreign’ by our immune system and likely depends on the context. As part of our genome, ERV protein

expression during thymocyte selection would contribute to negative selection of cognate T-cell receptors (TCRs) in the thymus and the establishment of immunological tolerance towards ERV-derived antigens. While many ERVs are epigenetically repressed and therefore not expressed, some are widely expressed. Indeed, these commonly-expressed endogenous retrotransposon-derived proteins contribute to 'self' peptides presented to developing T cells (Ebert et al., 2009; Lo et al., 2009; Young et al., 2012). Therefore, we would not expect to be immune tolerant to ERVs that are heavily DNA methylated and therefore not expressed in the thymus. This notion is supported by a study in which the immunisation of mice and non-human primates with endogenous retroelement antigens resulted in strong T-cell and B cell responses suggesting that some endogenous antigens are sensed as 'foreign' by the host (Kershaw et al., 2001; Sacha et al., 2012). Once the TCR and BCR repertoires are established in an individual, any endogenous antigens that were not expressed during negative selection in the thymus will likely to be seen as foreign. Therefore, ERVs must be examined individually with some potentially viewed as "self" and some as "foreign".

Importantly, when LTRs escape epigenetic repression, they can function as promoters and enhancers and potentially create neo-antigens through aberrant splicing (see section 1.2: The impact of ERVs on their hosts). For example, a HERV-E locus upstream of the human *CD5* gene has been shown to initiate the production of an alternatively spliced *CD5* mRNA (Renaudineau et al., 2005) and in turn a truncated intracellular form of the CD5 protein. Adaptive immunity against ERVs is not limited to T cell responses as antibody responses too have been found to be important (Young et al., 2012)

The best known example of a T cell population that is specific for an ERV is a particular TCR V β family that is reactive to endogenous MMTV-encoded superantigens (Ebert et al., 2009; Lo et al., 2009). These T cells are not negatively selected and interestingly, a fraction of the reactive thymocytes develop into regulatory T cells (T_{Reg}), which can suppress stimulation from MMTV while retaining their ability to react to self-antigens (MMTV). Superantigen-reactive T_{Reg} cells circulate in the periphery and prevent excessive immune pathology (Myers et al., 2013; Punkosdy et al., 2011).

Rationale

It is well known that embryonic stem cells possess intrinsic factors that repress retroviruses and part of this silencing activity is linked to the KZNF / KAP1 pathway (Rowe et al., 2010; Teich et al., 1977; Turelli et al., 2014; Wolf and Goff, 2007; Wolf and Goff, 2009; Wolf et al., 2008b). This mechanism has been documented in both mouse and human ES cells and is consistent with *in vivo* data showing that the activity of several ERVs peaks during early development before their KAP1-mediated repression around the blastocyst stage (Grow et al., 2015; Macfarlan et al., 2012; Rowe et al., 2010). KAP1 repression results in histone and DNA methylation, the latter which might replace the role of KAP1 in adult tissues, since KAP1 is no longer required to repress IAP elements in mouse embryonic fibroblasts (Rowe et al., 2010).

The focus of my thesis, therefore, is to investigate the role of KAP1 in retroviral repression outside of development. The relevance of this work is that retroviral transcriptional repression mechanisms operating in adult tissues are essential to prevent untoward innate and adaptive immune recognition of ERVs that may lead to autoimmunity. Moreover, ERV innate sensing has recently been shown to lie at the heart of cancer and immunity (Chiappinelli et al., 2015; Roulois et al., 2015).

Several lines of evidence suggest that KAP1 may restrict retroviruses beyond early development. Firstly, intriguingly, it has been shown that in mouse neural progenitor cells (NPCs), KAP1 still plays a role in repressing certain ERV families and still also influences transcriptional networks through ERVs (Brattas et al., 2017; Fasching et al., 2015) like in ES cells (Rowe et al., 2013a; Turelli et al., 2014).

However, since NPCs are isolated from developing mouse embryos, this work does not reveal a role for KAP1 beyond development. Secondly, studies in ZFP809 knockout mice showed an upregulation of VL30-pro retroviral transcripts in somatic tissues (Wolf et al., 2015). However, this was not a conditional knockout and it was molecularly linked back to the failure of these mice to establish VL30 repression and DNA methylation in early development due to the lack of Zfp809 protein. Thirdly, Setdb1 has been shown to repress ERVs in mouse B lymphocytes and this histone methyltransferase functions at least at some sites with KAP1 (Collins et al., 2015). Finally and most importantly, KAP1 represses several cellular genes through ERV binding sites in mouse liver (Ecco et al., 2016)

Interestingly, not all ERVs are silenced during development. It has been shown that some ERV loci act as tissue-specific enhancers and are not DNA methylated (Feuchter and Mager, 1990), for example in human T cells and fetal brain (Ling et al., 2002; Meissner et al., 2008). ERV expression that occurs in type II T-independent B cell responses also suggests that at least a subset of ERVs are not DNA methylated as otherwise they would need to undergo DNA demethylation prior to their activation (Zeng et al., 2014). These data imply that certain ERV loci are poised to become activated in response to stimuli or tissue-specific transcription factors. Whether KAP1 plays a role in maintaining ERVs in a poised state is unknown.

Aims

My project aims to explore two aspects of KAP1 repression of ERVs in differentiated human cells, the first focused on function and the second on biological relevance.

Aim 1. KAP1 repression

I sought to determine whether KAP1 represses ERVs in differentiated human cells using HERV-K14C as a model because it is KAP1 regulated at least in human ES cells (Turelli et al., 2014). HERV-K14C is a relatively recent integrant into the mammalian genome. It is unique to the Catarrhini lineage so it could potentially even regulate human-specific genes. However, given its young age and low copy number, its unlikely to regulate a network of genes. Of interest, HERVK14C is expressed at the 8-cell stage of development and also in the process of iPSCs production from CD34+ cells (Friedli et al., 2014). In addition, ENCODE data shows it to be KAP1 bound.

Specific objectives:

- Is KAP1 necessary to repress endogenous HERVK14C in differentiated human cells?
- Can KAP1 establish *de novo* repression of newly introduced HERVK14C?
- Is DNA methylation involved in the repression of HERVK14C?
- What other repeats and genes are regulated by KAP1 in differentiated cells?
- Does KAP1 repression proceed through the same mechanism in differentiated cells as in embryonic cells e.g. through SETDB1 and H3K9me3?

Aim 2. Biological relevance

The next question I aimed to focus on is whether KAP1 epigenetic repression prevents innate sensing of ERVs in differentiated cells? This is a cutting-edge topic in light of two recent articles that have revealed that the drug 5-Aza-cytidine, which is used to treat cancer patients likely functions by activating transcription of ERVs that are usually epigenetically silenced. This leads to their innate sensing and an anti-tumour response (Chiappinelli et al., 2015; Roulois et al., 2015). Epigenetic silencing of ERVs has been linked to DNA methylation in these studies and my aim is to determine whether KAP1 and associated epigenetic modifiers are necessary to sustain this DNA methylation block.

Specific objectives:

- Establish whether KAP1 could be involved in the 5-Aza-mediated effect by examining if 5-Aza treatment leads to the activation of KAP1-regulated ERVs as well as ISG induction.
- Deplete KAP1 and associated epigenetic candidates to see if this can phenocopy the effects of 5-Aza treatment on ISG induction.

Pursue the mechanism, for example assessing whether activation depends on loss of histone or DNA methylation or both, what the PAMPs are and whether ISG induction depends on RNA or DNA sensing or both.

Chapter 2: Materials and Methods

2.1 General solutions, buffers and materials

Table 1 List of commonly used materials and buffers.

Material/Buffer	Composition	Manufacturer (Cat. no.)
LB broth	20g/L H ₂ O	Sigma (L3022)
LB agar	37g/L H ₂ O	Millipore (1.10283.0500)
TAE buffer	1000x	AppliChem (A4686)
Agarose gel	1% in TAE (w/v), 4% EtBr (v/v)	Sigma (A9539)
DNA loading buffer	6x	Thermo Scientific (#R0611)
Gene Ruler 1kb	0.5ug/ul	Thermo Scientific (#SM0311)
NuPage LDS sample buffer	4x	Invitrogen (NP0007)
PageRuler Prestained	-	Thermo Scientific (26616)
Protein Ladder		
Running buffer	0.1% SDS, 1.44% Glycine, 0.3% Tris-base	-
Transfer buffer	1.44% Glycine, 0.3% Tris-base, 2% methanol	-
Wash buffer	0.1% Tween (v/v), PBS	

2.2 Western Blotting

2.2.1 SDS polyacrylamide gel electrophoresis (PAGE)

Table 2 Composition of SDS-PAGE used in the studies.

Solution	Composition
Stacking gel (6%)	20% Acrylamide, 8% Bis-acrylamide, 0.5% SDS, 125mM Tris pH6.8, 0.1% TEMED, 0.5% APS
Running gel (10%)	33.3% Acrylamide, 13.3% Bis-acrylamide, 0.5% SDS, 375mM Tris pH8.8, 0.03% TEMED, 0.5% APS

2.2.2 SDS-PAGE running

Approximately 2×10^6 cells were counted, washed with PBS and lysed in the sample buffer with 5% of β -mercaptoethanol. The samples were then sonicated at 20 Hertz (Hz) for 90 seconds and heated at 95°C for 5 minutes. Lysates were then loaded onto handcast SDS-PAGE gels in TRIS-GLYCINE running buffer using the mini-PROTEAN tanks (Biorad). The samples were stacked at 60V for 30 minutes and at 120V until a good separation of proteins of interest was achieved.

2.2.3 Transfer

Polyvinylidene difluoride (PVDF) membranes (Armesham, 10600023) were first immersed in methanol for 5 minutes and then transferred to the transfer buffer. The proteins on the gel were then transferred onto the PVDF membrane via wet transfer under 200V for 90 minutes on ice. For overnight transfer, 20V was used and the transfer was done in the cold room at 4°C .

2.2.4 Protein visualisation

Transferred membranes were blocked with 5% non-fat dried milk (w/v) (Sainsbury's) in wash buffer for 2 hours at room temperature (or overnight at 4°C). All antibodies were diluted in wash buffer with 3% non-fat dried milk (w/v) and the blocked membranes were incubated with the antibodies according to the manufacturer's protocols. Membranes were washed with wash buffer for 10 minutes and the washing was repeated three times after each antibody staining. For visualization, the membranes were incubated briefly with enzymatic

chemiluminescent reagents (Amersham, RPN2232, 2235, 2236) and exposed to High Performance chemiluminescence film (Amersham, 28906839) according to the manufacturer's instructions. Films were then developed via the film processor (Konica Minolta, SRX-101A).

2.2.5 Antibodies

Table 3 Antibodies used in the studies.

Target antigens	Manufacturer, Cat. no.	Comments (Assay)
Actin	Chemicon, MAB1501	Mouse monoclonal (WB)
KAP1	Millipore, MAB3662	Mouse monoclonal (WB)
PCNA	Millipore, clone PC10	Mouse monoclonal (WB)
OCT3/4	Santa Cruz, sc-5279	Mouse monoclonal (WB)
PE-Cy CD3 (ε-chain)	BD Bioscience, 555341	Mouse (FACS)
FITC-CD4	BD Bioscience, 345768	Mouse (FACS)
PE-OCT3/4	eBioscience, 12-5841-80	Rat (FACS)
FITC-IgG1 (κ isotype control)	BD Bioscience, 555909	Mouse (FACS)
PE-Cy5 IgG2a (κ isotype control)	BD Bioscience, 555575	Mouse (FACS)
PE-IgG2a (κ isotype control)	eBioscience, 12-4321-80	Rat (FACS)
MPP8	Proteintech, 16796-1-AP	Rabbit polyclonal (WB)
Mx1	Atlas, HPA030917	Rabbit polyclonal (WB)
BST2	Invitrogen, 16-3179-82	Mouse monoclonal (WB)
HERV-K Capsid	Ambio, HERM-1831-5	Mouse monoclonal (WB)
ZNF33A	Thermo Fisher, PA5-43694	Rabbit polyclonal (WB)
ZNF37A	Santa Cruz, sc-98283	Rabbit polyclonal (WB)
ZNF320	Abcam, ab105824	Rabbit polyclonal (WB)

2.3 Cell culture

Table 4 Cell types used throughout the studies.

Cell type	Medium condition	Growth condition
NTERA-2	DMEM, 10% FCS, 50µg/ml Pen/Strep	37°C, 5% CO ₂
HEK293T	DMEM, 10% FCS, 50µg/ml Pen/Strep	37°C, 5% CO ₂

HEK293	DMEM, 10% FCS, 50µg/ml Pen/Strep	37°C, 5% CO ₂
HeLa	DMEM, 10% FCS, 50µg/ml Pen/Strep	37°C, 5% CO ₂
CD4+T cells	RPMI, 10% HS, 50µg/ml Pen/Strep, 10U/ml IL-2	37°C, 5% CO ₂ (Activated by α-CD3, α-CD28 and 25U/ml IL-2)
PBMCs	RPMI, 20% FCS, 50µg/ml Pen/Strep, 10U/ml IL-2	37°C, 5% CO ₂ (Activated by 3ug/ml PHA and 10U/ml of IL-2)

References for the cells used were as follow:

NTERA-2 (Lee and Andrews, 1986); HEK293T (Phillips et al., 1999); HEK293 (Graham et al., 1977); HeLa (Scherer et al., 1953)

2.3.1 Antibiotics for cell culture

Table 5 Antibodies used for cell selection in the studies.

Antibiotic	Concentration	Manufacturer
Puromycin	2.5µg/ml	Sigma-Aldrich

2.3.2 Cell culture maintenance

NT2 cells were split every 3 days at a ratio of 1:3 via cell scraping and the cells were kept in T25 flasks. All other adherent cells were split every 3 to 4 days at a ratio of 1:4 using trypsin. THP-1 cells were kept in T75 flasks and split every 2 days at a ratio of 1:3. THP-1 cells were maintained at the density of 2×10^5 cells per ml. Both PBMCs and CD4+ T cells were activated under the conditions listed in Table 2.2 for 72 hours. Primary cells were maintained at a cell density of 1×10^6 cells per ml and were kept for not more than 14 days post activation.

2.3.3 Primary cell isolation

PBMCs were isolated from either a buffy coat or from healthy donors using Lymphoprep (Axis-shield, 1114544) according to the manufacturer's instructions. From PBMCs, CD4⁺ T cells were negatively selected and isolated using a CD4⁺ T cell isolation kit (Miltenyl Biotec,130-042-401) according to the manufacturer's instructions. The purity of the cells was verified using flow cytometry following antibody staining using antibodies listed in Table 3.

2.4 CRISPR/Cas9 genome editing

A pool of guide RNA (sgRNA) pairs each targeting different exons of the human KAP1 gene were designed using the website: <http://crispr.mit.edu/> and were ordered from Sigma-Aldrich. The oligos were resuspended in H₂O to yield a final concentration of 100μM. To anneal the sgRNAs, 1μl of each sense and antisense oligo was mixed with 1μl of T4 ligation buffer (reference 2.5.4), 0.5μl of T4 polynucleotide kinase (NEB, M0201S), and 6.5μl of H₂O. The mixture was then incubated at 37°C for 30 minutes, heated up to 95°C for 5 minutes and then cooled gradually on the bench for 90 minutes before checking the concentration. This oligo duplex was then cloned into the pSpCas9(BB)-2A-Puro (PX459) plasmid at the BbsI site. The digested backbone was run on an agarose gel and the right product was extracted using the QIAquick gel extraction kit (Qiagen, 28704).

For the ligation, 50ng of the cut backbone plasmid was mixed with 1ul of the oligo duplex (after 1:250 dilution), 5μl of 2X Quickligase buffer, 1μl of Quickligase (NEB, M2200L), and H₂O up to a final volume of 10μl. The ligation mix was left at room

temperature for 10 minutes before adding 1.5µl of 10x PlasmidSafe buffer, 1.5µl of 10mM ATP, and 1ul of PlasmidSafe exonuclease (Epicentre, E3101K). The mixture was incubated at 37°C for a further 30 minutes after which the ligation products were ready for HB101 transformation (of 5ul). The inserts were then verified through Sanger sequencing using the U6_forward primer.

Once verified, the plasmids were transfected into HeLa and 293T cells for 24 hours before puromycin was added for selection. After 24 hours of selection or until all the control cells were killed, the cells were then expanded and used for single cell cloning. Surviving cells were first diluted to the concentration of 0.5 cells per 100µl and plated into a 96-well-plate at the volume of 100µl per well. Once clones had grown, around ten were expanded and the presence or absence of KAP1 protein was determined by Western Blotting using the KAP1 antibody listed in Table 3.

2.5 Cloning

2.5.1 DNA preparation

All plasmid DNA used for this study was grown in HB101 *E. coli*. expanded in house and plasmids were isolated using Qiaprep Mini (Qiagen,, 27106) and Midi (Qiagen, 12145) kits according to the manufacturer's instructions and were resuspended in TE buffer.

2.5.2 Bacterial Transformation

Around 10ng of plasmid DNA was incubated with competent HB101 on ice for 20 minutes and then subjected to heat shock at 42°C for 40 seconds. Cells were

then returned on ice for a further 2 minutes before spreading them onto pre-warmed LB agar with the appropriate antibiotic.

Table 6 Antibiotic used for bacterial selection in the studies.

Antibiotic	Working concentration	Manufacturer
Ampicilin	100µg / ml	Calbiochem, 171254

2.5.3 Oligo annealing for shRNA cloning

The sequence of the target gene was downloaded from the Emsembl website: <http://www.ensembl.org/index.html> and the shRNA oligos were designed using the website: <http://bioinfo.clontech.com/rnaidesigner/>. 5'-phosphorylated oligos (sense and antisense strands) were ordered from Sigma-Aldrich and resuspended in water to a final concentration of 100µM. 10µl of each oligo pair were mixed with 2.5µl of 2M NaCl. The mixture was first heated up to 98°C for 5 minutes and then left to cool down to room temperature on the bench. After that, 350µl of H₂O, 40µl of 3M NaAc, and 1.1ml of EtOH were added to the mixture. The mixture was then transferred to -80°C overnight. After chilling, the mixture was spun down by centrifugation (Eppendorf, Centrifuge 5430 R) at 4°C at a maximum speed for 45 minutes. The pellet was then resuspended in 50µl of distilled H₂O was ready for ligation thereafter.

2.5.4 Ligation

Ligations were carried out using T4 DNA ligase (Thermo Fisher, EL0011) unless otherwise stated according to manufacturer's instruction. Briefly, inserts and

digested backbone were mixed (10ul final volume) and incubated at 16°C overnight or at room temperature for at least 2 hours. 4ul of the ligation product was used for transforming the HB101 (Section 2.5.2).

2.5.5 Sequencing

All clones were verified by Sanger sequencings provided by Genewiz (U.K.).

2.6 Lentiviral vectors

2.6.1 Vector Production

All lentiviral vectors used were VSV-pseudotyped. The lentiviral genome is produced in combination with VSV envelope to alter the vector tropism. Around 2×10^6 293T cells were seeded onto a 10cm plate. After 24 hours, cells were co-transfected with 1.5µg of the genome plasmid, 1µg of p8.91 and 1µg of pMDG2 encoding VSV-G. Fresh media was replaced 24 hours post-transfection and supernatant was harvested at 48 hours post-transfection. The vectors were used either neat or concentrated via ultracentrifugation (20,000g for 2 hours at 4°C). The ultracentrifugation was done as per the manufacturer's instructions using a Sorvall 100SE (Hitachi). Vectors were resuspended in ice cold PBS for storage at -80°C.

2.6.2 Virus titration

For adherent cells, different doses of GFP lentivectors were titrated onto 10^4 cells/well (2×10^4 cells per ml) in a 24-well plate. For suspension cells (CD4+ T cells

and PBMCs), cells are maintained at a minimum of 10^6 cells per ml for titration. After 24 hours post-transduction, the cells were washed and replenished with fresh media. After 48 hours, the cells were harvested for a GFP readout using flow cytometry. Within the linear range, virus doses that gave 5-20% of infection were then selected to calculate the infectious viral units per ml (IU/ml) using the formula: (Percentage of infection (%) * 10^4 cells)/ Volume of the vector added (ml). The calculated values were then used as a guide to obtain the same input dose across different vectors for the experiment.

For knockdown and overexpression experiments, around 500 μ l of non-concentrated vector or up to 2 μ l of concentrated vector was used to transduce 1×10^5 cells per well in 12 well plates. After 48 hours, puromycin was added for 24 hours or until all control cells died. Cells were then used at stated time-points following antibiotic selection for Western blot and qRT-PCR to verify the knockdown efficiency and for retrotransposon readouts.

2.7 Polymerase chain reaction (PCR)

2.7.1 PCR for Bisulfite sequencing

PCRs yielding products <500bp using bisulfite- converted DNA were conducted using the Hot Start Taq kit (Qiagen, 203203).

Reaction mix:

Table 7 Reaction mix for Hot start PCR.

	Volume (μ l)
Converted DNA	4
10x Buffer	5
DNTPs (10mM)	1
Primer mix (Forward and reverse 10 μ M)	1.25
Hot Start Taq polymerase	0.25
H ₂ O	38.5
Total	50

A generic cycling protocol for the PCR is shown below and the exact annealing temperature is noted in the 'comments' section for individual primer sets and the final extension time is dependent on the size of the amplicon.

Table 8 Cycling protocol for Hot start PCR.

Step	Hold	PCR cycle (35 cycles)			Final extension
		Denature	Anneal	Extension	
Temperature	95.0°C	95.0°C	58.0°C	72.0°C	72.0°C
Time	15 minutes	30 seconds	1 minute	1 minute	7 minutes

2.7.2 Phusion PCR

Phusion (NEB, M0530) PCR was used to amplify the genomic HERVK14C LTR sequence for cloning purposes and the reaction mix is as follows:

Table 9 Reaction mix for Phusion PCR.

	Volume (μl)
Template (Genomic DNA)	100ng
5x Phusion buffer	10
DNTPs (10mM)	1
Forward primer	2.5
Reverse primer	2.5
DMSO	1.5
Phusion polymerase	0.5
H ₂ O	Variable
Total	50

A generic cycling protocol for the PCR is shown below and the exact annealing temperature is noted in the 'comments' section for individual primer sets and the final extension time is dependent on the size of the amplicon.

Table 10 Cycling protocol for Phusion PCR.

Step	Hold	PCR cycle (30 cycles)			Final extension
		Denature	Anneal	Extension	
Temperature	98.0°C	98.0°C	60.0°C	72.0°C	72.0°C
Time	15 minutes	10 seconds	30 seconds	40 seconds	7 minutes

2.8 HERVK14C LTR reporter

2.8.1 Cloning

A HERVK14C that is bound by KAP1 was first identified using ENCODE binding data and its LTR sequence along with or without the following PBS sequence was then amplified through PCR using Phusion High Fidelity DNA polymerase (New England Biolabs, M0530L). Amplicons were then digested with XhoI and BamHI and ligated with digested pPGK-GFP. Clones were then verified using Sanger sequencing.

2.8.2 Reporter experimental set-up

Lentiviral vectors were produced as described in 2.6.1 using the genome plasmids with the right inserts in KAP1 knockout 293T cells. The infectious titer of the vectors was then determined by titrating the vectors on KAP1 knockout 293T cells and a GFP readout was determined using flow cytometry.

2.9 Flow cytometry

All GFP readouts were determined by flow cytometry using either a BD Accuri C6 machine or a FACS Calibur machine both from BD Bioscience. Analysis was performed using FlowJo and live cells were gated on using the forward scatter height (FSC-H) and side scatter height (SSC-H) parameters, which correspond to cell size and granularity, respectively. 10,000 live cells were used for analysis in each sample. A non-transduced (or control antibody treated) cell sample was used to define the percentage of GFP, or OCT-4 positive cells, as measured using the FL1 or FL2

channels respectively.

2.9.1 Intracellular staining

1×10^6 cells per condition were trypsinized, counted, and washed with media. The cells were then resuspended in residual media and fixed by resuspending in 100 μ l of formaldehyde-based fixation buffer (eBioscience). Samples were then incubated at room temperature in the dark for 30-60 minutes. After incubation, 2ml of permeabilization buffer (eBioscience) was added directly, followed by centrifugation at 300-400g for 5 minutes at room temperature. This wash was repeated before resuspending pellets in 100 μ l of permeabilization buffer for staining. The recommended amounts of fluorochrome-labeled antibodies were then added and incubated for 20-60 minutes at room temperature in the dark. After this, 2 washes with permeabilization buffer were performed and then pellets were resuspended in PBS and samples acquired on a flow cytometer.

2.9.2 Buffers

Commercial buffers were used for the intracellular staining (eBioscience, 00-5523-00).

Table 11. Buffers used for intracellular staining

Buffer	Comment
Permeabilisation buffer	00-8333-56
Fixation/ permeabilisation diluent	00-5223-56
Fixation/ permeabilisation concentrate	00-5123-43

2.10 Luciferase assay

2.10.1 Transfection

On day 1, 293T cells are plated at a concentration 1×10^5 cells/ ml in a 24 well plate (with 500ul per well). The transfection mix was prepared as follows:

Table 12 Transfection mix for luciferase based assay.

	Volume per well
Reporter plasmid	X
Thymidine-kinase renilla plasmid	X/10
Opti-MEM	25 μ l
Fugene	1.5 μ l

Different amounts of the reporter plasmids were titrated onto the cells in order to ensure that the luciferase readout was within the detection limit of the luminometer. The renilla plasmid was added at a 1 to 10 ratio to the luciferase reporter plasmid. Fugene was added last before a 20 minute incubation at room temperature and then the mix was added drop-wise to cells. Sendai virus was added 24 hours before harvesting cells as a positive control ISG-reporter induction as shown in Figure 39.

2.10.2 Luciferase readout

The cells were harvested 48 hours post-transfection for the luciferase readout using the Dual –Luciferase[®] Report Assay System kit (Promega, E1960) and the Glomax 96 microplate luminometer (Promega). Briefly, the cells were washed with cold PBS and 100 μ l of 1x passive lysis buffer was then added to each well. Samples were then subjected to a round of freeze-thaw treatment. After thorough

resuspension, 20µl of the thawed samples was subsequently transferred to a white 96 well plate and 30ul of the luciferase reagent was added to each well. The signal was then detected in the illuminometer using the standard 'Steady Glow' protocol by Promega. After multiple readings, 30ul of Renilla substrate was then added directly into each well before the renilla readings were made.

An uninfected negative control was used to measure the background signal and this value was subtracted from all other readings. The output firefly luciferase relative light units (RLU) were then normalised to the corresponding renilla RLU. The fold induction of the luciferase signal was determined by calculating the fold change between the normalised RLU in knockdown cells and empty vector transduced cells.

2.11 RNA extraction

Cells were first trypsinized, spun down and washed with PBS once. The cell pellet was then lysed and RNA was harvested using an RNeasy mini kit (Qiagen, 74104) according to the manufacturer's instructions. The RNA was resuspended in RNase-free water.

2.12 Quantitative reverse transcription polymerase chain reaction (qRT-PCR)

2.12.1 cDNA synthesis

RNA was first incubated with Turbo DNase (Thermo Fisher Scientific, am1907) at 37°C for 25 minutes. DNase inactivating beads were then added to the

mixture, mixed and left at room temperature for 3 minutes. The treated RNA was then spun down and the supernatant was collected. The concentration of RNA was determined using the Nanodrop 1000 (Thermo Scientific).

500ng of purified RNA was mixed with 1µl of dNTPs (10mM) and random primers (150ng/µl) (Invitrogen, 58875) and incubated at 65°C for 5 minutes followed by incubation on ice for a further 5 minutes. dTTs (Thermo Fisher Scientific, 18064-014), RNase OUT (Thermo Fisher Scientific, 10777-019) and SuperScript II reverse transcriptase (Thermo Fisher Scientific, 18064-014) were then added together with the buffer provided. The mixture was incubated at 42°C for 50 minutes followed by 70°C for 15 minutes.

2.12.2 Reaction mix

Reaction mix:

Table 13 Reaction mix for qRT-PCR.

	Volume per reaction (µl)
Fast SYBR Green master mix	12.5
Forward and reverse primer mix	2.0
cDNA template	2.0
H ₂ O	8.5
Total	25.0

Constituents for the reaction mix:

Table 14 Reagents used in qRT-PCR reaction.

Components	Manufacturer, Cat. No.	Comment
Fast SYBR Green master mix	Thermo Fisher Scientific, 4385612	-
Primer mix	Sigma-Aldrich	470µl of H ₂ O, 15µl of forward primer (100µM), 15µl of reverse primer (100µM)
cDNA template	-	Diluted (e.g. 1:50) in H ₂ O

2.12.3 Cycling protocol

The protocol for the qRT-PCR is as follow:

Table 15 Cycling protocol used for qRT-PCR.

Step	Hold	PCR cycle (40 cycles)	
		Denature	Anneal / Extension
Temperature	95.0°C	95.0°C	60.0°C
Time	10 minutes	15 seconds	1 minute

2.13 Taqman qPCR

The DNA template was first extracted as described in 2.16 and the reaction mix is set up as follows:

Table 16 Reaction mix used for Taqman qPCR.

	Volume per reaction (μl)
Taqman Universal master mix	12.5
Forward and reverse primer mix	2.0
Probe	1.5
H ₂ O	7.0
DNA template	2.0
Total	25.0

A DNA template made from a single copy GFP integrant 293T cell line was used as a reference to quantify the copy number of integrants in samples tested.

Table 17 Reagents used in Taqman qPCR reaction.

Components	Manufacturer, Cat. No.	Comment
Taqman Universal master mix	Life Technology, 4440040	-
Primer mix	Sigma-Aldrich	Stock concentration of 3μM
Probe	Sigma-Aldrich	Stock concentration of 250 nM
DNA template	-	Diluted to stock concentration of 50ng/μl

The protocol for Taqman qPCR is as follow:

Table 18 Cycling protocol used for Taqman qPCR.

Step	Hold	PCR cycle (40 cycles)	
		Denature	Anneal / Extension
Temperature	95.0°C	95.0°C	60.0°C
Time	10 minutes	15 seconds	1 minute

2.14 RNA sequencing

Total RNA samples were diluted to 50ng/μl (and 1μg was supplied per sample) for RNA sequencing, which was performed by the UCL Genomics Platform.

2.14.1 Library preparation

Samples were processed using Illuminas Truseq Stranded mRNA LT sample preparation kit (Illumina, RS-122-2101) according to manufacturer's instructions.

Deviations from the protocol were as follows:

250ng total RNA was used as starting material and fragmented for 10 minutes and 14 cycles of PCR were used.

Briefly, mRNA was isolated from total RNA using oligo dT beads to pull down poly-adenylated transcripts. The purified mRNA was fragmented using chemical fragmentation (heat and divalent metal cation) and primed with random hexamers. Strand-specific first strand cDNA was generated using Reverse Transcriptase and Actinomycin D. This allows RNA dependent synthesis while preventing spurious DNA-dependent synthesis. The second cDNA strand was synthesized using dUTP instead of dTTP, to mark the second strand. The cDNA is then "A-tailed" at the 3' end to

prevent self-ligation.

Full-length TruSeq adaptors, containing a T overhang are ligated to the A-Tailed cDNA. These adaptors contain sequences that allow the libraries to be uniquely identified by way of a 6bp Index sequence. Successfully ligated fragments were enriched with 14 cycles of PCR. The polymerase is unable to read through uracil, so only the first strand is amplified.

2.14.2 Sequencing

Libraries to be multiplexed in the same run are pooled in equimolar quantities, calculated from Qubit and Bioanalyser fragment analysis. Samples were sequenced on the NextSeq 500 instrument (Illumina, San Diego, US) using a 43bp paired-end run resulting in >15million reads per sample.

2.14.3 Data analysis (Genes)

Run data were demultiplexed and converted to fastq files using Illumina's bcl2fastq Conversion Software v2.16. Fastq files were then aligned to the human genome NCBI build 37.2 using Tophat 2.014 then deduplicated using Picard Tools 1.79. Reads per transcript were counted using HTSeq and differential expression was estimated using the BioConductor package DESeq2.

2.14.4 Data analysis (Repeats)

TrimGalore v0.4.0 with Cutadapt v1.6 was used to remove adaptors and trim

read ends using default parameters (quality phred score cutoff =20, minimum required sequence length for both reads before a sequence pair gets removed = 20 bp and m trimming error rate: 0.1) on the mRNA-seq data. The number of sequence pairs removed varied from 2.11% to 11.21%. Reads were first mapped against the human genome hg38 using Bowtie2 v2.2.4 with default parameters, which means that:

- A) It searches for multiple alignments and reports the best one;
- B) Reads are not excluded based on mismatches, which is an appropriate approach for mapping genes and repetitive regions and;
- C) Reads were not filtered based on mapping quality nor duplicated reads removed.

Reads were subsequently mapped against the RepBase v20.06 human library using Bowtie2v2.2.4 as above. The samtools v.1.19 idxstat utility was used to extract the number of mapped reads per repeat, that were inputted into the R package DESeq (<https://bioconductor.org/packages/3.2/bioc/html/DESeq.html>) to identify differentially expressed repeats between KAP1 KOs and controls. DESeq uses a test based on the negative binomial distribution and a shrinkage estimator for the distribution's variance to find differentially expressed repeats between conditions. P-values were adjusted for multiple testing with the Benjamini-Hochberg false-discovery-rate (FDR) procedure.

2.15 Chip-sequencing analysis

Chromatin-immunoprecipitation-sequencing data for human ESCs and 293T cells were downloaded from NCBI Gene Expression Omnibus (GEO) under accession numbers GSE57989 (HuESC) and GSE27929 (293T). TrimGalore v0.4.0 was used to remove adaptors and trim read ends, and reads were mapped against the human genome (hg19 assembly) using Bowtie2 v2.2.4. Peaks were called in each replicate using Macs2 v 2.1.1, and the bioconductor package DiffBind (<https://bioconductor.org/packages/DiffBind>) was used to construct Venn diagrams and identify overlapping ESC and 293T peaks. Human repeat and gene locations were downloaded from the UCSC browser (RepeatMasker and RefGene tables) and the repeats and genes closest to the overlapping ESC/293T peaks were identified using bedtools-2.17.0. Chip-sequencing correlations were analysed using the Chip-Cor website: http://ccg.vital-it.ch/chipseq/chip_cor.php.

2.16 DNA extraction

Cells were first trypsinized, spun down and washed with PBS once. The cell pellet was then lysed and DNA was harvested using a DNeasy Blood & tissue kit (Qiagen, 69504) according to the manufacturer's instructions.

2.17 DNA methylation analysis

1µg of DNA was used for bisulphite conversion using the EpiTect Bisulfite Kit (Qiagen, 59104) following the manufacturer's protocol. 4 µl of converted DNA was then amplified through PCR using HotStarTaq DNA polymerase (Qiagen, 203203)

(see Section 2.7.1). Primers for this procedure were designed using the site: <http://urogene.org/methprimer/> and the PCR products were then cloned into the pCR2.1TOPO vector using a TOPO cloning kit (Invitrogen, 45-0641). A T7P primer was used to sequence the products. The DNA methylation status of the clones was measured and analysed using the QUMA online tool: <http://quma.cdb.riken.jp> by the Riken Institute.

2.18 Statistical analysis

All data in the studies are presented as standard deviation (where there are three or more samples) assessed by using two-tailed, unpaired Student t tests (see individual figure legends for further details). A P-value of ≤ 0.05 was considered as statistically significant (****: $p \leq 0.0001$, ***: $p \leq 0.001$, **: $p \leq 0.01$, *: $p \leq 0.05$).

2.19 Oligonucleotides

Table 19 Primers for qRT-PCR in the studies.

Primer	Sequence (5'-3')	Comment (No. of loci detected)
B2M	F: TGC TCG CGC TAC TCT CTC TTT R: TCT GCT GGA TGA CGT GAG TAA AC	qRT-PCR (1)
GAPDH	F: GGG AAA CTG TGG CGT GAT R: GGA GGA GTG GGT GTC GCT GTT	qRT-PCR (1)

KAP1	F: AAG GAC CAT ACT GTG CGC TCT AC R: ACG TTG CAA TAG ACA GTA CGT TCA C	qRT-PCR (1)
LIPA4_1	F: TCA CCA ATA TCC GCT GTT CTG R: GTC TGT TGG AGT TTA CTG GAG G	qRT-PCR (63)
LIPA4_2	F: TGA TAC CCA GGC AAA CAG G R: TCT AAC AGT CAG GAC CCT CAG	qRT-PCR (1025)
HERVK14C_1	F: AAY AGC ACT GGA GCC CTT R: CGA CTG TGA TGG TTS AYT TTG	qRT-PCR (5)
HERVK14C_2	F: GTA ATT GTG AGT ACC CAA AAT CTC R: ACC TTG TCC CAA TCT TTT AC	qRT-PCR (5)
SVA D VNTR	F: GCT GCC CAT CGT CTG AG R: TCC TCA CCT CCC AGA CAG	qRT-PCR (7)
ISG56	F: CCT GAA AGG CCA GAA TGA GG R: TCC ACC TTG TCC AGG TAA GT	qRT-PCR (1)
CCL5	F: CCC AGC AGT CGT CTT TGT CA R: TCC CGA ACC CAT TTC TTC TCT	qRT-PCR (1*)
CXCL10	F: GTG GCA TTC AAG GAG TAC CTC R: GCC TTC GAT TCT GGA TTC AGA	qRT-PCR (1*)

Mx α	F: ATC CTG GGA TTT TGG GGC TT R: CCG CTT GTC GCT GGT GTC G	qRT-PCR (1*)
I κ B α	F: CTC CGA GAC TTT CGA GGA AAT R: GCC ATT GTA GTT GGT AGC CTT	qRT-PCR (1)
HKR1	F: CCA AAA CTC ATT GCT CAG CTG R: GAG AAA ATC AGA GGG CAG GAG	qRT-PCR (1*)
ZNF45	F: CTG TAC CGA GAT GTG ATG CTG R: TCT GGG TTG CCA TCT TCA TC	qRT-PCR (1)
ZNF274	F: CTG AAG ATG GAA GCC TGA GTG R: TGT CCT TAT AAC GGA ACT GCC	qRT-PCR (1)
ZNF300	F: GGG TAT CCA GTT TCC AAA CCA G R: GTC TCC CAT CTG CCT GAT ATT C	qRT-PCR (1)
ZNF607	F: GGA CAT TCC GTA TCT AAG CCA G R: CTC TCC TCT TGT TTC TTC CCT C	qRT-PCR (1)
ZNF140	F: CTG GTC TCA CTG GGT CTT TC R: TTC ACT TCC CTT TTC CCC AG	qRT-PCR (1*)

ZNF350	F: AGA AAA TGA TCC AGG CCC AG R: AGG TTG CTG TAG TTC TCC AAC	qRT-PCR (1*)
ZNF432	F: CAA AGA CAG AGT ATC CCA GGT G R: CCG GTA CAA ATC CTT CTG AAA AG	qRT-PCR (1*)
MPP8	F: TGC CTG TAT CTG CCC AAA C R: CCT TGT CAG AAT CAT GCC TTT TC	qRT-PCR
Tasor	F: TGA AGA CAT TGC AGG TTT CAT TC R: CAT CCA GGC TAT CAA CAC CAG	qRT-PCR
Periphilin	F: GGT TCC AGT GTC AGT AGC AG R: TCA TTC TGC CGT TTG TAG GAG	qRT-PCR
SETDB1	F: CAA GCT GGG ACT ACA ATA CCG R: TCT GGT CTT TTG GAG TTC TGC	qRT-PCR
HERVK14C (Old primer)	F: CCA TTG TGC TCC ATT GGA AG R: CCC TCT GTG CCG ATT GAA AG	qRT-PCR (1)

The copy number of loci detected by the primers are determined using the in-silico PCR programme (<https://genome.ucsc.edu/cgi-bin/hgPcr>). * - Denotes cross-exonic primers. F: forward; R: reverse.

Table 20 shRNAs for knockdown in the studies.

shRNA	Sequence (5'-3')	Comment
shKAP1_1	F:	Targeting

	GAT CCG TAA GCA CAG GTT TGG TCT CAG TTC AAG AGA CTG AGA CCA AAC CTG TGC TTA TTT TTT ACG CGT G R: AAT TCA CGC GTA AAA AAT AAG CAC AGG TTT GGT CTC AGT CTC TTG AAC TGA GAC CAA ACC TGT GCT TAC G	coding regions
shKAP1_2	F: GAT CCG TAA GAA CTG GTA CTG GTG GTC TTC AAG AGA GAC CAC CAG TAC CAG TTC TTA TTT TTT ACG CGT G R: AAT TCA CGC GTA AAA AAT AAG AAC TGG TAC TGG TGG TCT CTC TTG AAG ACC ACC AGT ACC AGT TCT TAC G	Targeting coding regions
shKAP1	F: GAT CCG CCT GGC TCT GTT CTC TGT CCT TTC AAG AGA AGG ACA GAG AAC AGA GCC AGG TTT TTT ACG CGT G R: AAT TCA CGC GTA AAA AAC CTG GCT CTG TTC TCT GTC CTT CTC TTG AAA GGA CAG AGA ACA GAG CCA GGC G	Targeting UTR
shTasor	F: GAT CCG AGG AAG CTT GAG GAT CTA TTC AAG AGA TAG ATC CTC AAG CTT CCT CTT TTT TG R: AAT TCA AAA AAG AGG AAG CTT GAG GAT CTA TCT CTT GAA TAG ATC CTC AAG CTT CCT CG	Targeting coding regions
shMPP8	F: GAT CCA AGA AGA CCC CGA GAA AGG TTC AAG AGA CCT TTC TCG GGG TCT TCT TTT TTT TG R: AAT TCA AAA AAA AGA AGA CCC CGA GAA AGG TCT CTT GAA CCT TTC TCG GGG TCT TCT TGG ATC G	Targeting coding regions
shPeriphilin	F: GAT CCA GCT AAC CAC TCG CTC TAA TTC AAG AGA TTA GAG CGA GTG GTT AGC TTT TTT T R: GAA TTC AAA AAA AGC TAA CCA CTC GCT CTA ATC TCT TGA ATT AGA GCG AGT GGT TAG CTG GAT CG	Targeting coding regions

shSETDB1	F: GAT CCA CCA GAT GAT CCC TTC CAA TTC AAG AGA TTG GAA GGG ATC ATC TGG TTT TTT TAC GCG TG R: AAT TCA CGC GTA AAA AAC CAG ATG ATC CC TTC CAA TCT CTT GAA TTG GAA GGG ATC ATC TGG TG	Targeting coding regions
----------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------

F: forward; R: reverse.

Table 21 sgRNAs for generating CRISPR knockout cells.

sgRNA	Sequence (5'-3')	Comment
KAP1_ Ex1	F: CAC CGG AGC GCT TTT CGC CGC CAG R: AAA CCT GGC GGC GAA AAG CGC TCC	Targeting Exon 1
KAP1_ Ex9	F: CAC CGC GTC CTG GCA CTA ACT CAA C R: AAA CGT TGA GTT AGT GCC AGG ACG C	Targeting Exon 9

F: forward; R: reverse.

Table 22 PCR primers used for bisulfite sequencing.

Bisulfite PCR	Sequence (5'-3')	Comment
HERVK14C_LTR	F: AGG TTT AGG AAG GTG GAT TAT TTG R: ACA ACC CAA AAC TCT CAA ACT CTA C	Annealing temperature: 55.0°C
OCT-4	F: ATT TGT TTT TTG GGT AGT TAA AGG T R: CCA ACT ATC TTC ATC TTA ATA ACA TCC	Annealing temperature: 58.0°C
SVA	F: TTG TAA TTT TTT TGT TTG ATT TTT TTG T R: TAC ACT CCA ACC TAA ACA CCA TTA A	Annealing temperature: 58.0°C

LIPA4	F: GGG TGA TTT TTG TAT TTT TAA TTG AG R: AAA AAA AAT TCC CTA ACC CCT TAC	Annealing temperature: 58.0°C
-------	--------------------------------------------------------------------------------------	-------------------------------------

F: forward; R: reverse.

Table 23 Sequencing primers used throughout the studies.

Sequencing primer	Sequence (5'-3')	Comment
T7	TAA TAC GAC TCA CTA TAG GG	TOPO cloning
CPPT_F	CAG GCC CGA AGG AAT AGA AG	Reporter cloning
U6_F	GAG GGC CTA TTT CCC ATG ATT CC	CRISPR cloning
AF159	ATG CCA ATT GCT CCT TCT CTA GGC GCC GGA AT	shRNA cloning

F: forward; R: reverse.

Table 24 Primers and probes used for Taqman qPCR.

Taqman qPCR	Sequence (5'-3')	Comment
GFP	F: CTG CTG CCC GAC AAC CAC R: ACC ATG TGA TCG CGC TTC TC Probe: [6FAM] CCA GTC CGC CCC TGA GCA AAG ACC [BHQ1]	
Human Albumin	F: GCT GTT CAT CTT GTG GGC TGT R: ACT CAT GGG AGC TGC TGG TTC Probe: [6FAM] CCT GTC ATG CCC ACA CAA ATC TCT CC [BHQ1]	

F: forward; R: reverse.

Table 25 Primers used for amplifying genomic sequence in reporter plasmid generation.

Reporter PCR	Sequence (5'-3')	Comment
Con_LTR	F:	Contain

	TCG AGG CAC GCG TTC TGG CGC CCA ACG TGG GGC TCC CCA TAA TCC R: TCG AGG ATT ATG GGG AGC CCC ACG TTG GGC GCC AGA ACG CGT GCC	consensus PBS-K and are annealed before cloning upstream of the LTR promoter in the reporter plasmid.
Chr15_LTR	F: GCC TCG AGG GAT TAT GGG GAG CCC CAT GTT GGG CGC CAG AGC ACG CGT TGT GAG AAA GAG AGT TTC TGA GGT GC R: GCG GAT CCT GTT GGG GAA ACC AGC CC	Forward primer contains specific PBS-K sequence in antisense orientation.
LTR	F: GCC TCG AGA GAC CTG TGA GAA AGA GAG TTT CTG AGG TGC R: GCG GAT CCT GTT GGG GAA ACC AGC CC	Annealing temperature: 57.0°C

F: forward; R: reverse.

2.20 Plasmid list

Table 26 List of plasmids used in the studies.

Plasmid	Comment
p8.91Ex	Packaging plasmid for lentivector production
pMD.G	Envelope plasmid for lentivector production
pKAP1	Wild type KAP1 overexpression plasmid
pPGK-GFP	Lentiviral vector-based GFP reporter
pSIREN.HIV	Lentiviral vector for shRNA expression
pzfp809	Zfp809 expressing plasmid
pzfp819	Zfp819 expressing plasmid
pPro-GFP	GFP expression plasmid with PBS-Pro upstream of the promoter
SVA-SV40-Luc	Luciferase plasmid driven by SV40 minimal promoter with SVA VNTR sequence upstream of the promoter
LIPA4-SV40-Luc	Luciferase plasmid driven by SV40 minimal promoter with LIPA4 sequence upstream of the promoter
pZNF91	ZNF91 expressing plasmid
pZNF93	ZNF93 expressing plasmid

pTK-Renilla	Renilla plasmid driven by thymidine kinase promoter
pISGs-Firefly-Luc	Luciferase plasmid driven by various ISG promoters

2.21 Drug list

Table 27 List of drugs used in the studies.

Drug	Working concentration	Manufacturer
5-Aza-2'-deoxycytidine	7 μ M	Sigma-Aldrich
Trichostatin A	200nM	Sigma-Aldrich

2.22 Data access

mRNA-sequencing data are being submitted to the NCBI Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) database. Accession numbers for the public data are as follows: 293T: GSE27929 (Chip-seq), HuESC: GSE57989 (Chip-seq), 293T: GSE44267 (mRNA-seq), HeLa: this study (mRNA-seq), Macrophages: GSE36952 (mRNA-seq), CD4+ T cells: GSE69549 (mRNA-seq).

Chapter 3: KAP1 regulates ERVs in differentiated human cells

3.1 HERVK14C as a model to study the role of KAP1 in differentiated cells

KAP1-mediated repression of ERVs in ESCs has been well documented by multiple studies (Maksakova et al., 2013; Rowe et al., 2010; Rowe et al., 2013b; Turelli et al., 2014). However, it is unclear if KAP1 is necessary to suppress ERVs outside of development. Importantly, one study showed that several murine ERVs become derepressed in mouse liver following the liver-specific knockout of KAP1 (Ecco et al., 2016) but to date there are no publications about the role of KAP1 in differentiated human cells. To dissect the relationship between KAP1 and ERVs in differentiated human cells, we first selected HERVK14C, because it is an endogenous retrovirus known to be repressed by KAP1 in human ESCs (Friedli et al., 2014; Turelli et al., 2014). We found HERVK14C to be restricted to Great Apes and Old World Monkeys through sequence analysis (Figure 13A) and it is a low copy number ERV making it relatively easy to study (Figure 13B). The number of proviral copies depicted (Figure 13B) include only those that have internal regions (at least one open reading frame), which we chose to focus on because KAP1 is known to target several internal sites such as the primer binding site (PBS) and group specific antigen (GAG) (Sadic et al., 2015; Schlesinger et al., 2013; Wang et al., 2014a; Wolf and Goff, 2007; Wolf et al., 2008b). These copies were also flanked by LTRs at both 5' and 3' ends (Figure 13B). While solo LTRs are not included in our initial qRT-PCR studies here, all types of repeats are included in our mRNA-sequencing analyses (Figure 18).

First, we depleted KAP1 using short hairpin RNA (shRNA), encoded by a VSVG-pseudotyped lentivector and using this system, we achieved efficient

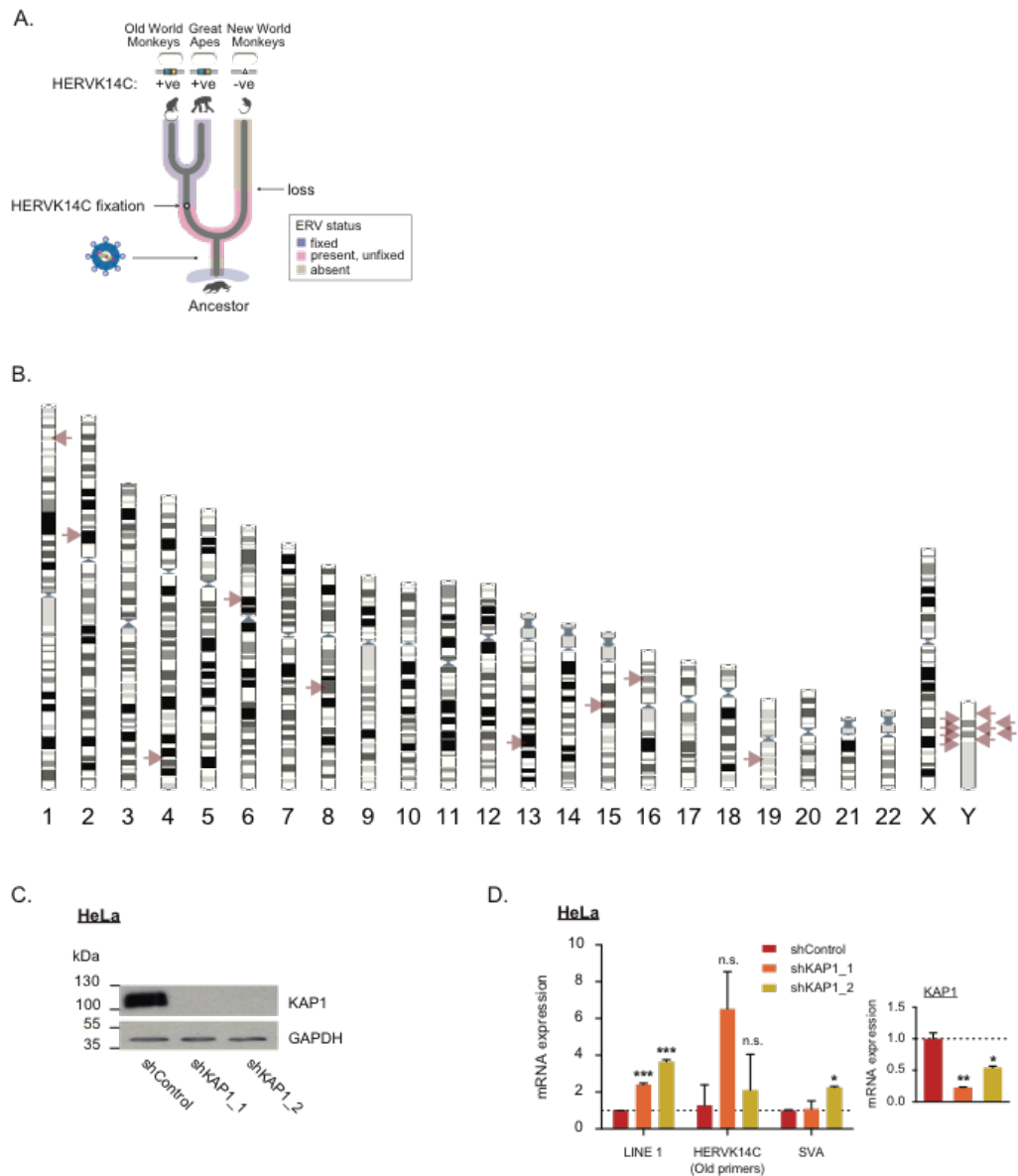


Figure 13 HERVK14C as a model to explore the role of KAP1.

(A) Schematic diagram depicting the age of HERVK14C.

(B) Chromosome map showing the loci of HERVK14C with at least one internal open reading frame flanked by both 5' and 3' LTRs.

(C) Western blot showing the knockdown efficiency of KAP1 using two distinct hairpin vectors.

(D) Plot showing the expression of LINE1, HERVK14C and SVAs upon KAP1 knockdown. Here, the old primers were used to detect HERVK14C. The inset shows the level of KAP1 mRNA upon knockdown. One representative experiment of 2 experiments is shown here. All mRNA expressions were normalised to B2M.

Two-tailed unpaired Student t-tests were performed.

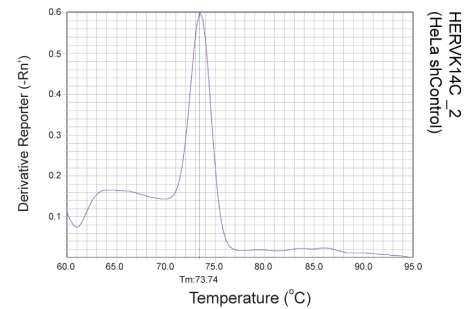
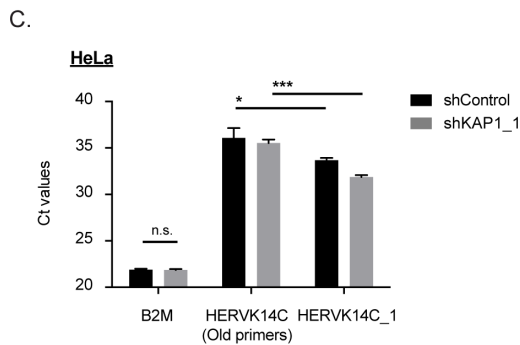
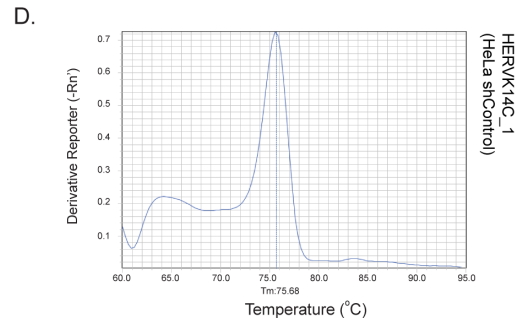
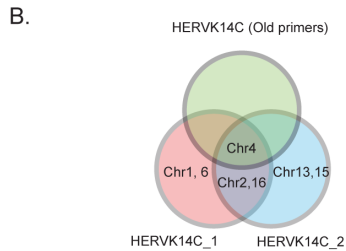
[illegible]

Figure 14 Design of new HERVK14C primer sets.

- (A) Diagram with highlighted sequences in blue showing the region upon which the new primers were designed. Point mutations within the blue region are annotated with asterisks.
- (B) Loci targeted by the old and new qRT-PCR primers. The new primer sets are labelled as 'HERVK14C_1' and 'HERVK14C_2'.
- (C) Histogram showing the comparison of qRT-PCR Ct values using the old and new HERVK14C primer sets. The same template was used in all the reaction mixes.
- (D) Diagrams showing the melting curves of the new primer sets. T_m: melting temperature. Two-tailed unpaired Student t-tests were performed.

knockdown of KAP1 in HeLa cells (Figure 13C). We checked the expression of HERVK14C as well as of global LINE1 and SVA elements using qRT-PCR. LINE1 and SVA elements were only modestly upregulated, while HERVK14C was more potently derepressed (up to 6 fold) (Figure 13D). However, results for LINE1 and SVAs were significant, whereas HERVK14C upregulation was not significant due to its extremely low and variable expression level at baseline (Figure 13D). Indeed, the HERVK14C primers detect only a single copy of HERVK14C on Chromosome 4 (Turelli et al., 2014). To circumvent this issue, we designed two new primer sets to recognize multiple copies based on a HERVK14C alignment (Figure 14A & B). We then verified the new primer sets in terms of their Ct values and specificity (Figure 14C & D). The new primer sets produce significantly lower Ct values following qRT-PCR on both control and knockdown samples, as expected (Figure 14C). This improvement did not compromise their specificity as shown by their melting curves (Figure 14D).

3.2 KAP1 mediates HERVK14C repression in undifferentiated NTERA-2 cells

As part of the validation of our model, we utilised the pluripotent human embryonic carcinoma NTERA-2 cell line to confirm KAP1 repression of HERVK14C. Oct4 is a marker for undifferentiated embryonic cells so we started by staining NTERA-2 cells for intracellular Oct4 protein (also known as POU5F1) to confirm the undifferentiated status of these cells. We found that the vast majority of NTERA-2 cells expressed Oct4, while this marker was undetectable in 293T cells, as expected (Figure 15A). We then depleted KAP1 in NTERA-2 cells and verified the loss of KAP1 protein by Western blot (day 6 post transduction) (Figure 15B). Oct4 expression, in

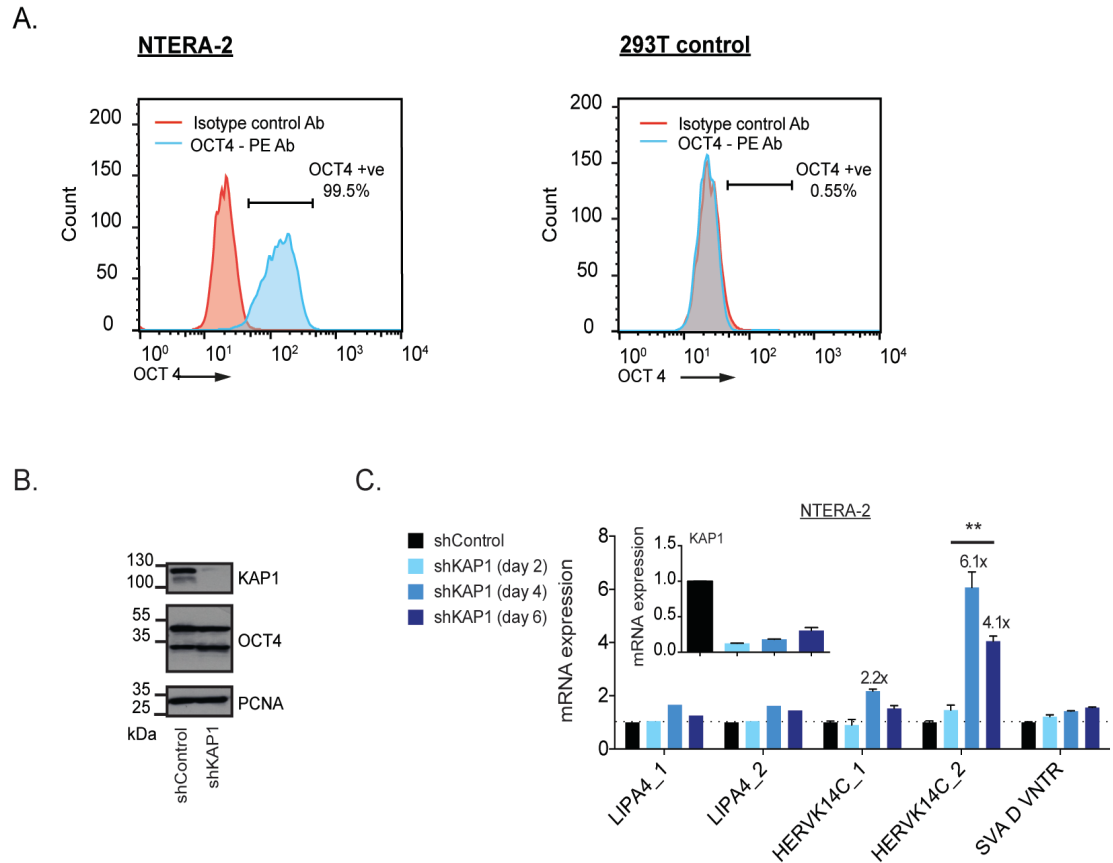


Figure 15 KAP1 regulates HERVK14C in NTERA-2 cells.

(A) Intracellular staining of OCT4 protein in NTERA-2 (left) and 293T control cells (right).
 (B) Western blot showing the level of KAP1 and OCT4 protein 6 days post-KAP1 knockdown.
 (C) Plots showing the expression of various repeats at multiple time points post-KAP1 knockdown. The inset shows the level of KAP1 mRNA. A representative experiment of 2 experiments is shown here. B2M was used as the normalising gene for mRNA expression. Data for all time points were pooled for statistical analyses. A two-tailed unpaired Student T-test was done and the p-value for HERVK14C_2 was 0.0023.

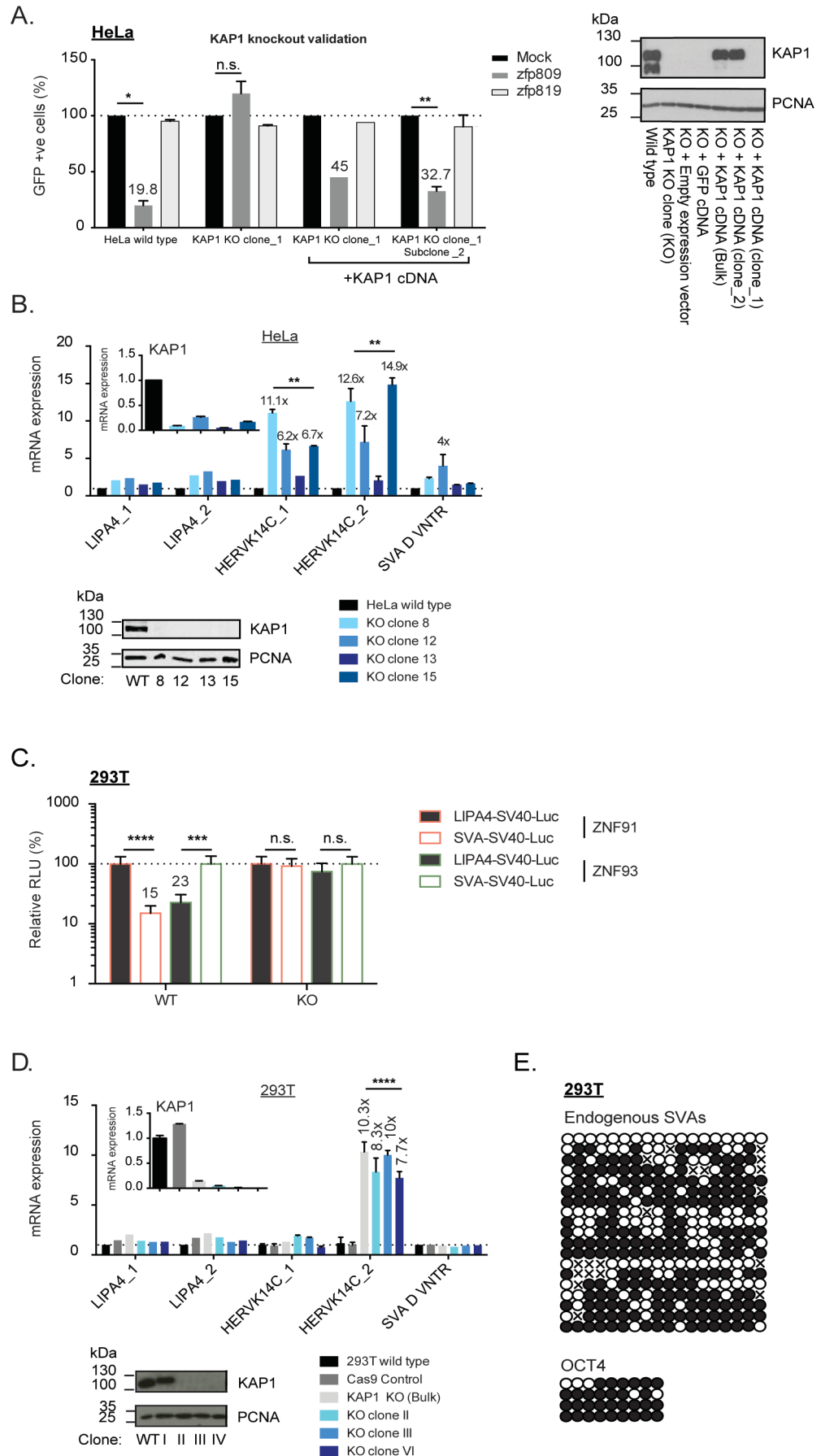


Figure 16 KAP1 represses HERVK14C in differentiated cells.

- (A) The bar chart shows the functional verification of KAP1 KO cells (left), n=2. The Western blot shows the expression of KAP1 in WT, KO and KAP1 reconstituted cells (right).
- (B) Graph showing the relative expression level of various repeats in WT and KO HeLa clones. The inlay shows the level of KAP1 mRNA while the Western blot (below) shows the protein level in the tested clones. B2M was used as the normalising gene. Data across all clones was pooled for statistical analyses. Clones 8, 12 and 15 were subsequently used for mRNA sequencing.
- (C) KAP1 KO 293T cells were functionally verified using a luciferase based readout assay.
- (D) Graph showing the expression level of various repeats in WT and KO 293T clones. The inlay shows the level of KAP1 mRNA while the Western blot (below) shows the protein level of the tested clones. B2M was used as the normalising gene. Data were pooled across all clones for statistical analyses.
- (E) Bisulfite sequencing of multiple loci of endogenous SVA D VNTRs and the OCT4 gene body. Each circle represents a CpG dinucleotide. Open circles represent a lack of methylation while filled circles represent methylation of the CpG.
- Two-tailed unpaired Student T-tests were done.

comparison was not affected in the same timeframe (Figure 15B). HERVK14C upregulation could be observed by as early as day 2 post knockdown and this phenotype was sustained at least up to day 6 post knockdown (Figure 15C). We also looked at expression of LIPA4 and SVA D VNTR, both of which have previously been shown to be bound by KAP1 (Jacobs et al., 2014). However, KAP1-depletion had little or no effect on the expression of these elements compared to its effect on HERVK14C (Figure 15C). Of note, the Ct values of these retrotransposons were much lower than of HERVK14C in control NTERA-2 cells pointing to their higher expression level (and copy number) at baseline. Indeed, LINE1 elements have been previously documented to be readily expressed in NTERA-2 cells (Chen et al., 2012).

3.3 KAP1 represses HERVK14C in differentiated cells

We then asked the question of whether KAP1 was necessary to repress these retrotransposons in differentiated cells. We therefore repeated these knockdown experiments in 293T cells and HeLa cells and took advantage of the CRISPR/Cas9 genome-editing system to additionally generate KAP1 knockout differentiated cell lines. We first validated the HeLa KAP1 knockout clones functionally using a previously described zfp809 and Pro-PBS system (Wolf and Goff, 2009) using reporters generated before (Rowe et al., 2013a) (Figure 16A). In wild type cells, GFP expression was significantly reduced in the presence of zfp809, whereas in KAP1 KO clones (which were validated to be KAP1 KO by Western blot) (Figure 16A, right), such repression was completely abrogated, as expected. Zfp819 was used as a negative control. Upon complementation with KAP1 cDNA, GFP repression was rescued to a level similar to that induced in wild type cells (Figure 16A). We then

selected several KAP1 knockout clones and looked at the impact of knockout on retrotransposon expression (Figure 16B). All clones showed a consistent upregulation of HERVK14C compared to controls as detected using both primer sets, while SVA D VNTR elements were also upregulated in some of the clones tested (Figure 16B). We validated KAP1 knockout in these clones at both the mRNA and protein level (Figure 16B). Similarly, we verified KAP1 loss functionally in 293T cell knockout clones by using previously described KAP1 reporters (Jacobs et al., 2014). The SV40-Luciferase (SVA-Luc) reporter plasmids, which contained either an LIPA4 or SVA D VNTR (SVA) KAP1 target sequence were co-transfected with either ZNF91 or ZNF93 as previously described (Jacobs et al., 2014). ZNF91 mediated repression of SVA-SV40-Luc and ZNF93 mediated repression of LIPA4-SV40-Luc and both reporters were dependent on KAP1 (Figure 16C). We then assessed repeat expression in these clones and found that HERVK14C was upregulated in KAP1 KO 293T cells but this phenotype was restricted to the second primer set (HERVK14C_2) suggesting some heterogeneity in HERVK14C regulation between cell types (Figure 16D). Through bisulfite sequencing, we found high levels of DNA methylation at SVA element promoters (Figure 16E). This is consistent with previous findings that SVA elements are enriched for cytosine methylation (Turelli et al., 2014) and may explain the lack of SVA upregulation upon KAP1 knockout. The Oct4 gene body served as a positive control for endogenous DNA methylation (Figure 16E).

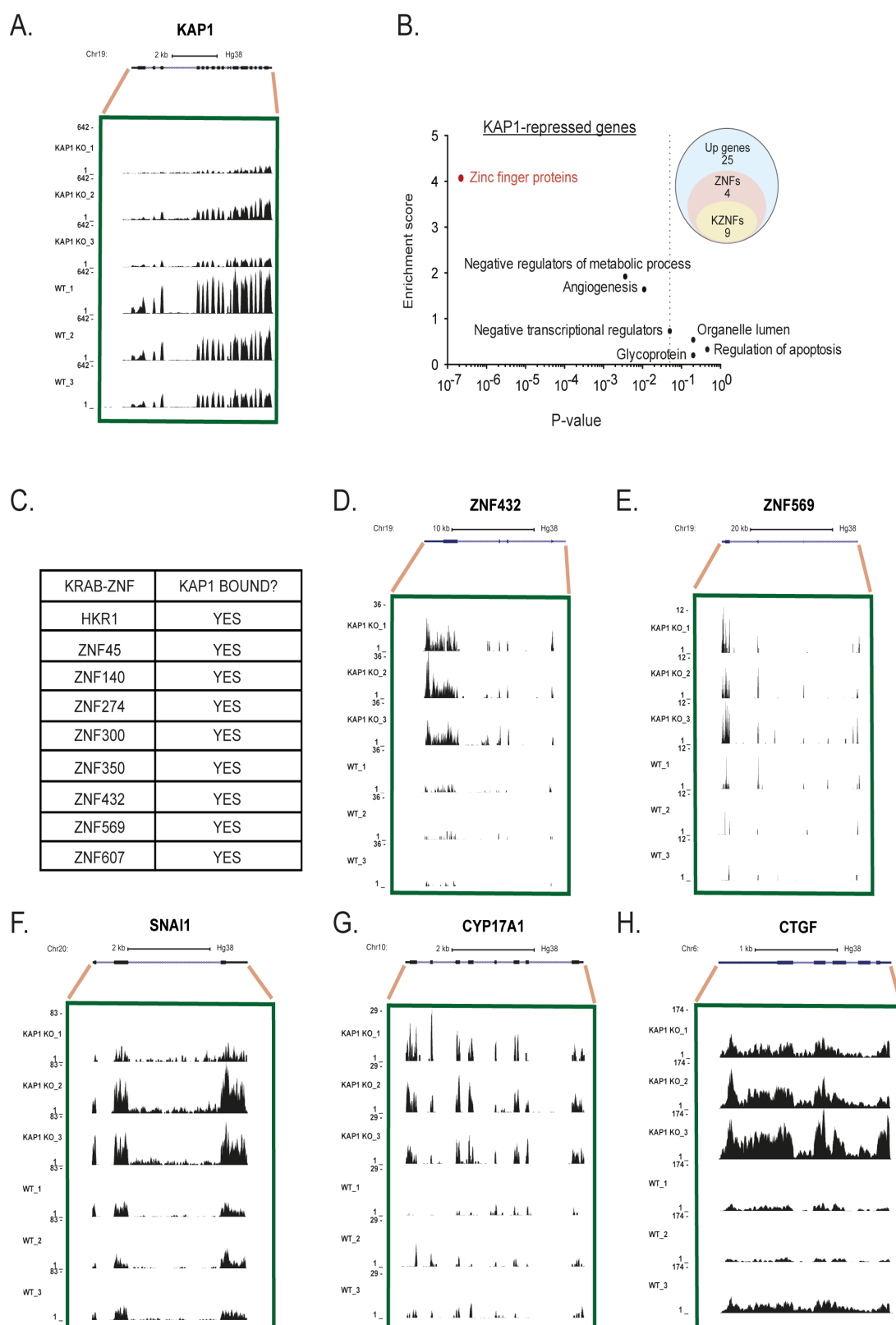


Figure 17 Analysis of KAP1 repressed cellular genes.

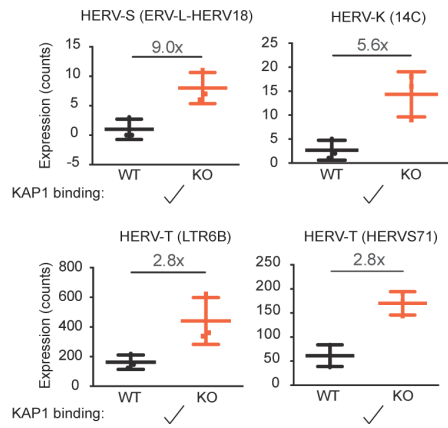
(A) Screenshots of UCSC genome browser tracks showing the mRNA level of KAP1. The upper three tracks illustrate the level of expression in KO samples while the lower three tracks show the levels in WT cells.

(B) Dot plot showing the enriched functional clusters among the upregulated genes (>2 fold with adjusted p-values < 0.05) in KAP1 KO cells compared to the WT according to DAVID gene ontology

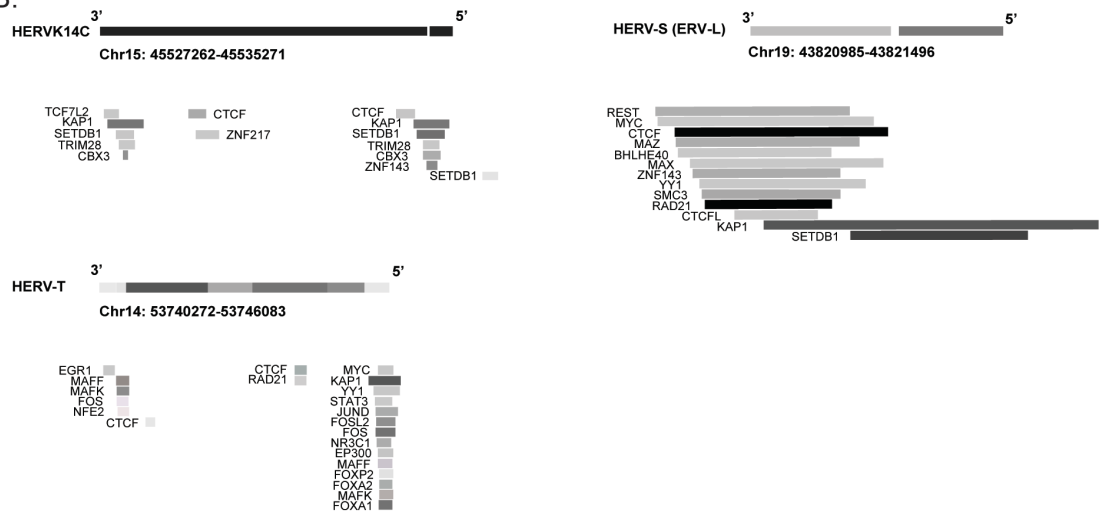
analysis. Inlaid Venn diagrams show the proportion of ZNFs and KZNFs among all the upregulated genes.

- (C) Table showing the identities of all the upregulated KRAB-ZNFs and the status of KAP1 binding to these loci according to ENCODE data.
- (D) – (H) Screenshots of UCSC genome browser tracks showing the mRNA expression level of various genes using our mRNA sequencing data. The upper three tracks illustrate the expression level in KO samples while the lower three tracks show the expression level in WT cells.

A.



B.



C.

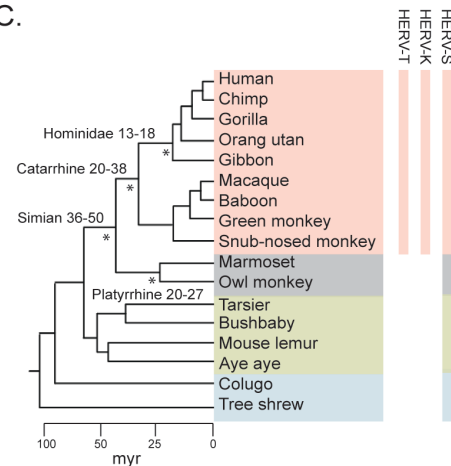


Figure 18 Analysis of HERVs repressed by KAP1.

- (A) Box plots (mean \pm SD) showing the difference in expression of upregulated repeats between WT and KO cells. The KAP1 binding status of these families is represented below the respective graphs.
- (B) Examples of loci in (A) bound by KAP1 and other transcription factors according to ENCODE data.
- (C) The estimated age of HERVs in (A) mapped onto the evolutionary tree. The estimated ages of the stated lineages are shown and marked with an asterisk. myr: million years.

3.4 KAP1 binds to and represses cellular genes

In order to further dissect KAP1 targets in differentiated cells in a systematic and unbiased manner, we employed mRNA sequencing using three wild type and three KAP1 knockout HeLa samples (KO clones 8, 12 and 15, see Figure 16B). Consistent with our qRT-PCR results, we saw a reduction in reads mapping the KAP1 transcript in all the knockout clones compared to the wild type ones (Figure 17A). In addition to KAP1, we also found several hundred cellular genes that were either positively or negatively affected following the knockout. When we focused on the upregulated genes (>2 fold with adjusted p values of ≤ 0.05), interestingly, DAVID analysis revealed zinc finger proteins (ZNFs) to be the class of genes most significantly affected (Figure 17B). Strikingly, the majority of these ZNFs (9 out of 13) have a KRAB domain (Figure 17B). These loci are all also direct KAP1-binding targets according to ENCODE data (Figure 17C), which is consistent with the reported role of KAP1 targeting to ZNFs, including through ZNF274 and ZNF75D (Fietze et al., 2010; Imbeault et al., 2017; Iyengar et al., 2011). Together, these observations reveal a functional role for KAP1 in fine-tuning cellular gene expression at multiple targets across chromosomes in differentiated human cells (Figure 17D-H). Indeed, this is consistent with the described role of KAP1 in regulating gene expression in several mouse tissues and cell types including T cells, B cells and mouse liver (Bojkowska et al., 2012; Santoni de Sio et al., 2012a; Santoni de Sio et al., 2012b).

3.5 KAP1 represses multiple HERV families in differentiated cells

We then mapped the mRNA sequencing reads to Repbase extending our analysis to repetitive elements and selected those that were >2 affected with p-

values of ≤ 0.05 . In addition to HERVK14C, we identified two other classes of ERVs (HERV-T and HERV-S) that were upregulated following KAP1 knockout (Figure 18A). As expected, these ERV families were also bound by KAP1 according to ENCODE data (Figure 18A & B). Interestingly, while HERV-T and HERV-K are both relatively recent, HERV-S appears to be an ancient lineage and present in diverse primates and other mammals (Figure 18C), suggesting it may be regulated through a conserved ZNF. In addition to KAP1 binding, we also found clusters of transcription factors binding to these ERVs suggesting that they may function as regulatory hubs and that repression may be dynamic (Figure 18B).

3.6 KAP1 knockout leads to downregulation of certain cellular genes and repeats

mRNA-sequencing also revealed several cellular genes to become downregulated (>2 fold with adjusted p-values of ≤ 0.05) following KAP1 knockout (Figure 19). Of particular interest, innate immune response genes were among the downregulated genes. Several publications link repetitive elements to innate immune responses (Chiappinelli et al., 2015; Roulois et al., 2015) and we will follow up this link in Chapter 5. Since KAP1 has also been described as a transcriptional activator (Chang et al., 1998; Rambaud et al., 2009; Singh et al., 2015), one possibility is that KAP1 promotes transcription of these genes in a wild type context. However, this hypothesis is not supported by the fact that KAP1 binding is not enriched in these gene clusters (Figure 21B). This suggests that these genes are indirect targets affected as a secondary consequence of KAP1 knockout. Likewise, we also found several classes of repetitive elements (HUERS-P1, LIPBA1_5, LIM2C_5) to

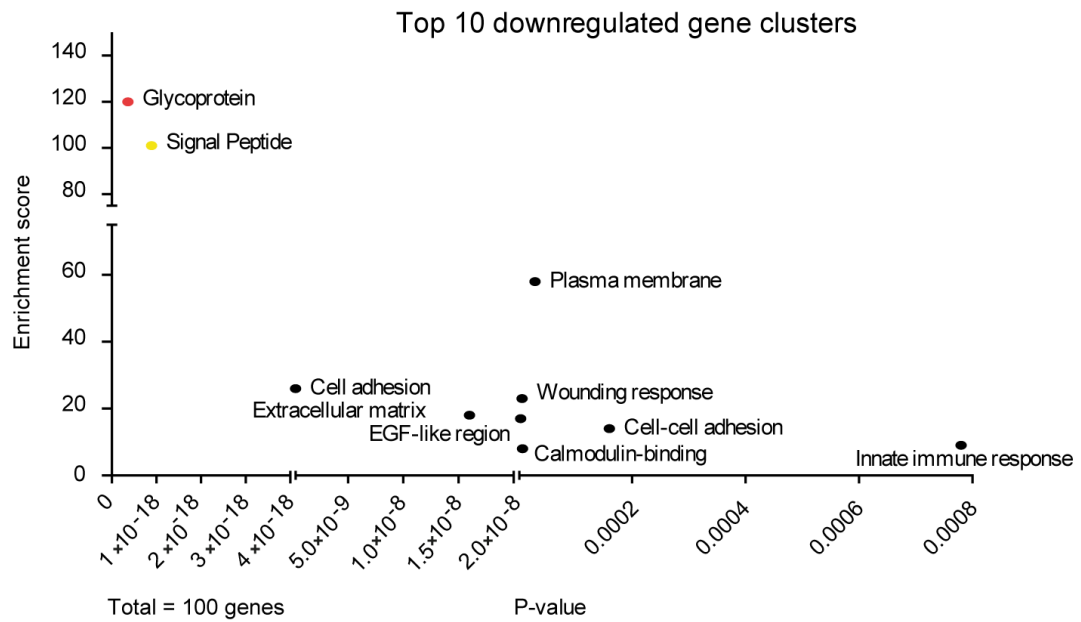


Figure 19 Analysis of cellular genes downregulated upon KAP1 KO.

Dot plot showing the enriched functional clusters of the downregulated genes in KAP1 KO cells compared to the WT according to DAVID gene ontology analysis.

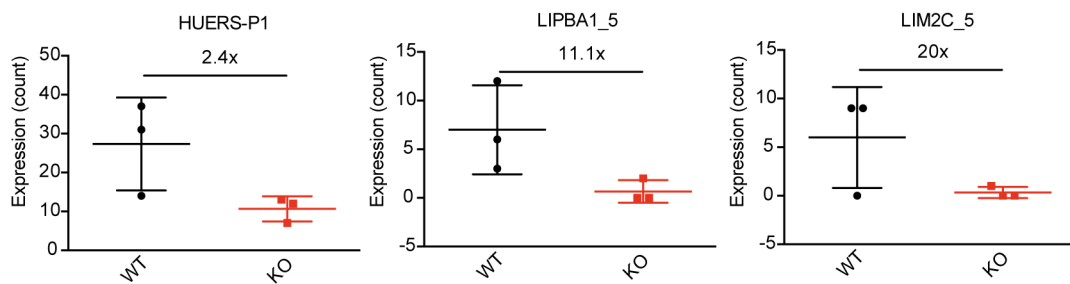


Figure 20 Analysis of repeats downregulated upon KAP1 KO.

Box plots (mean ± SD) showing the difference in expression of downregulated repeats between WT and KO cells.

be downregulated in our mRNA-sequencing reads (>2 fold with adjusted p-values of ≤ 0.05 , Figure 20).

3.7 KAP1 binds to common ERVs and ZNFs in undifferentiated and differentiated cells

Since we observed KAP1 repression of ERVs in NTERA-2 cells and differentiated cell lines, we hypothesised that KAP1 binds to common targets in both undifferentiated and differentiated cells. To test this hypothesis, we utilised public KAP1 ChIP-seq and ENCODE data from 293T cells and human ESCs. We identified 614 common KAP1 binding ChIP-seq peaks between the two cell types (Figure 21A). We then selected the nearest genes to these sites and interestingly, gene ontology analysis showed that the most enriched gene cluster was ZNFs (including 40 KZNFs out of the 61 ZNFs) mirroring our functional mRNA-sequencing data (Figure 17B). When we looked at where these loci fall in the genome, we found them to be highly enriched within LINE1 elements and ERV families compared to the abundance of these repeats in the genome (Figure 21C). We then looked at the distribution of LINE1 and ERVs and found a slight enrichment for ERV1 (ERV class I) elements within the common binding sites compared to the expected distribution of this ERV class (Figure 21D) as well as to its prevalence within either the ESC or 293T datasets. ERVs can generally be categorised into Class I, II or III based on their sequence similarities to their exogenous counterparts. Class I ERVs are Gammaretrovirus-like, Class II ERVs are Betaretrovirus-like while Class III are Spumavirus-like. Interestingly, the Class I clade includes HERV-T elements and this is consistent with KAP1 regulation and binding of this family (Figure 18A and B). Overall, these data show that a proportion

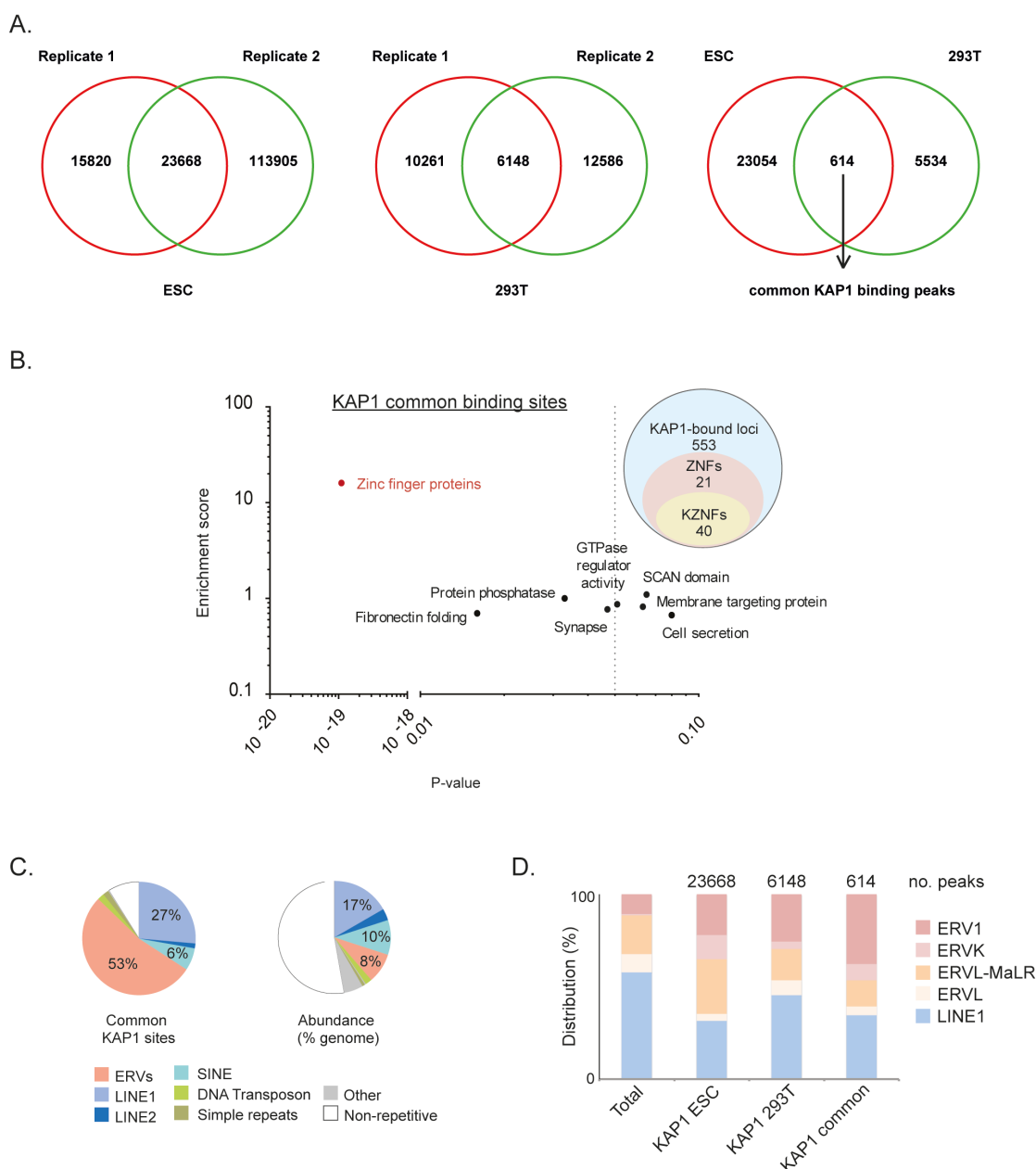


Figure 21 Analysis of conserved KAP1 binding loci as revealed through ChIP-seq.

- (A) Venn diagram showing the overlapping KAP1 binding sites between replicates of ESC and 293T cells and between the two cell types.
- (B) Dot plot showing the functional clusters enriched from the 614 KAP1 binding peaks in (A) using DAVID analysis. The Inlaid Venn diagram shows the proportion of ZNFs and KZNFs.
- (C) Pie charts illustrating the identity of common KAP1 binding sites (left) compared to the distribution of these features in the human genome (right).
- (D) Bar chart showing the distribution of LINE1s and ERV families in the genome (Total), and in the ESC KAP1 binding data (KAP1 ESC) and the 293T KAP1 binding data (KAP1 293T) and the common binding sites (KAP1 common).

of the KAP1 binding landscape is preserved from undifferentiated to differentiated cells.

3.8 KAP1 common binding sites are enriched for repressive chromatin marks

We then further hypothesised that KAP1 repression at these common sites is a conserved feature across multiple cell types. To address this, we probed the ENCODE data of epigenetic marks in various differentiated cells including primary cells to assess whether silent chromatin marks overlap with KAP1 common binding sites.

We indeed found that these loci correlate with KAP1 and SETDB1 binding as well as with H3K9me3 in multiple cell types (Figure 22A). In contrast, we detected no enrichment of the active chromatin marks, H3K4me1 and H3K27ac, at these common binding sites, as expected (Figure 22A). Since KAP1 is recruited to its targets by KZNFs (Iyengar et al., 2011), we hypothesised that a proportion of KZNFs must be widely expressed and at a level sufficient to recruit KAP1 to its targets in differentiated cells. Indeed, our HeLa mRNA-sequencing data revealed many KZNFs to be expressed at a low level (several examples are shown in Figure 17D & E). To investigate the profile of targets bound by KZNFs that are expressed in differentiated cells, we first identified the top one hundred most highly expressed KZNFs in 293T cells using public mRNA-sequencing data (Zuin et al., 2014). We then verified that these KZNFs are expressed at the protein expression level using the human protein atlas (Uhlen et al., 2010) and recorded their targets where known (Imbeault et al., 2017; Schmitges et al., 2016). Interestingly, many of these KZNFs bind to ERVs

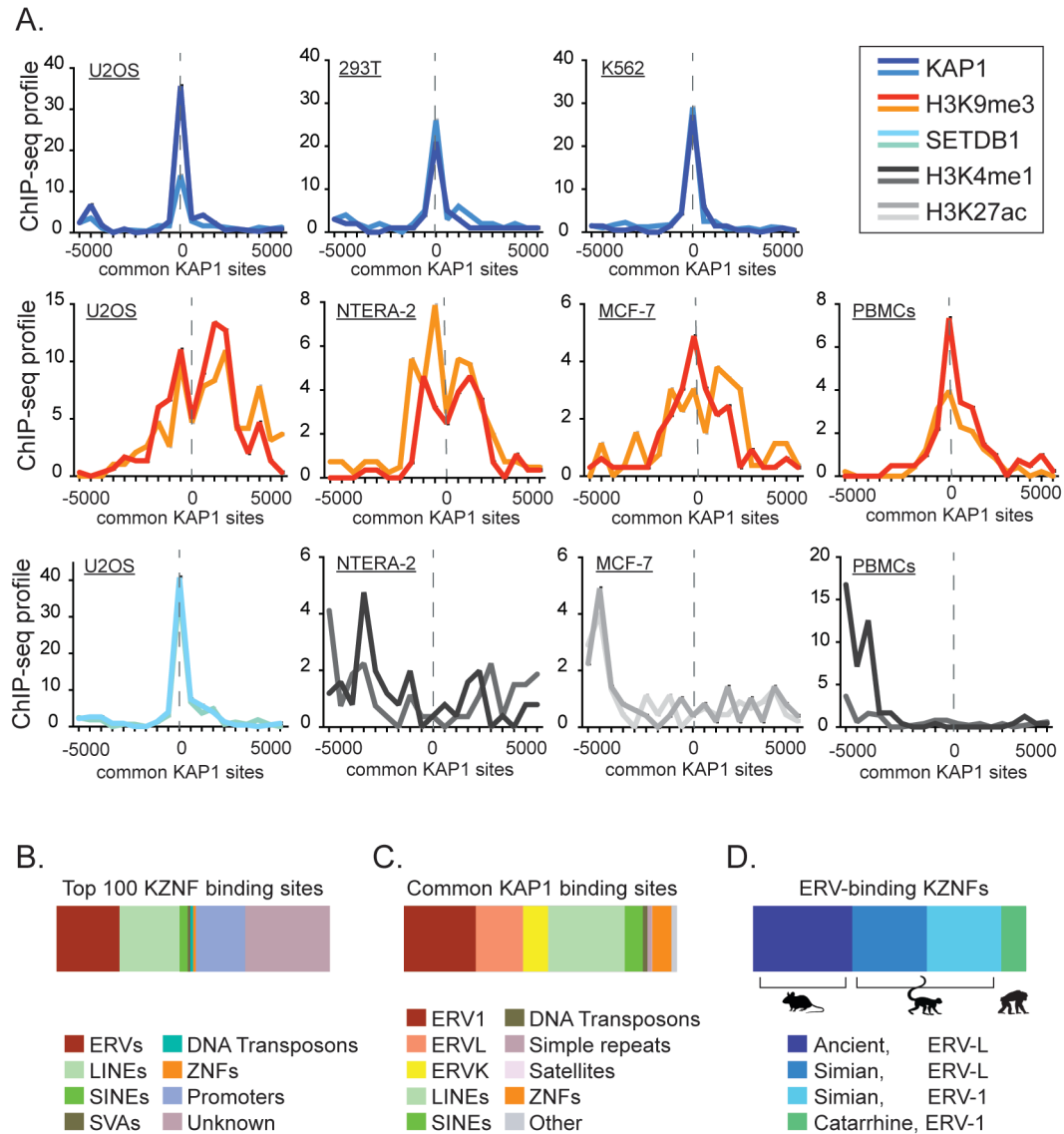


Figure 22 KAP1 common binding sites are enriched for repressive histone marks and overlap with KZNF-bound ERV loci.

- (A) ChIP-Cor analysis of the common KAP1 binding sites defined above (Figure 21A). Genomic coordinates were used as the input and intersected with ChIP-seq peaks using ChIP-Cor (Ambrosini et al., 2016). Each line plot shows two biological duplicates of ChIP-seq experiments from ENCODE.
- (B) Diagram showing the known targets within the genome of the top 100 most highly expressed KZNFs in 293T (as defined by RNA-seq RPKM values). The level of protein expression of these candidates was verified using the site: <http://www.proteinatlas.org/cell>.
- (C) Diagram showing the common KAP1 binding sites within the genome.
- (D) Distribution of the age and types of ERVs bound by the KZNFs in (B).

(Figure 22B), reflecting KAP1 common binding loci (Figure 22C). Intriguingly, the ERV-binding KZNFs are conserved among different primates and mammals reflecting the wide age spectrum of their ERV targets (Figure 22D). These data indicate that KAP1 and KZNFs regulate a subset of common ERVs and cellular genes between undifferentiated and differentiated cells and potentially also between species.

3.9 KAP1 mediates de novo repression of incoming ERVs through the primer binding site sequence

Next, since KAP1 is known to target a lysine primer binding site (PBS) sequence (Turelli et al., 2014; Wolf and Goff, 2007; Wolf et al., 2008b), we asked if KZNFs and KAP1 were expressed at sufficient levels in differentiated cells to induce *de novo* repression of incoming HERVK sequences. We first identified a KAP1-bound and repressed HERVK14C integrant on Chromosome 15 and used its LTR to construct a GFP-based reporter system to measure *de novo* KAP1 repression in differentiated cells. We then cloned its PBS sequence (PBSChr15), which was similar to a lysine PBS, upstream of the LTR in an antisense orientation (termed PBSChr15-LTR-GFP) and compared this vector to one containing the consensus lysine_{1,2} PBS (PBS) (termed PBS-LTR-GFP) and a control vector containing no PBS sequence (LTR-GFP) (Figure 23A). Using this system, we could demonstrate that the HERVK14C PBS sequence and the consensus PBS sequence were sufficient to induce modest GFP repression and this was dependent on KAP1 in both undifferentiated and differentiated cells (Figure 23B and C). We confirmed that the difference in GFP levels between wild type and knockout cells was not due to a difference in vector integration (Figure 23D).

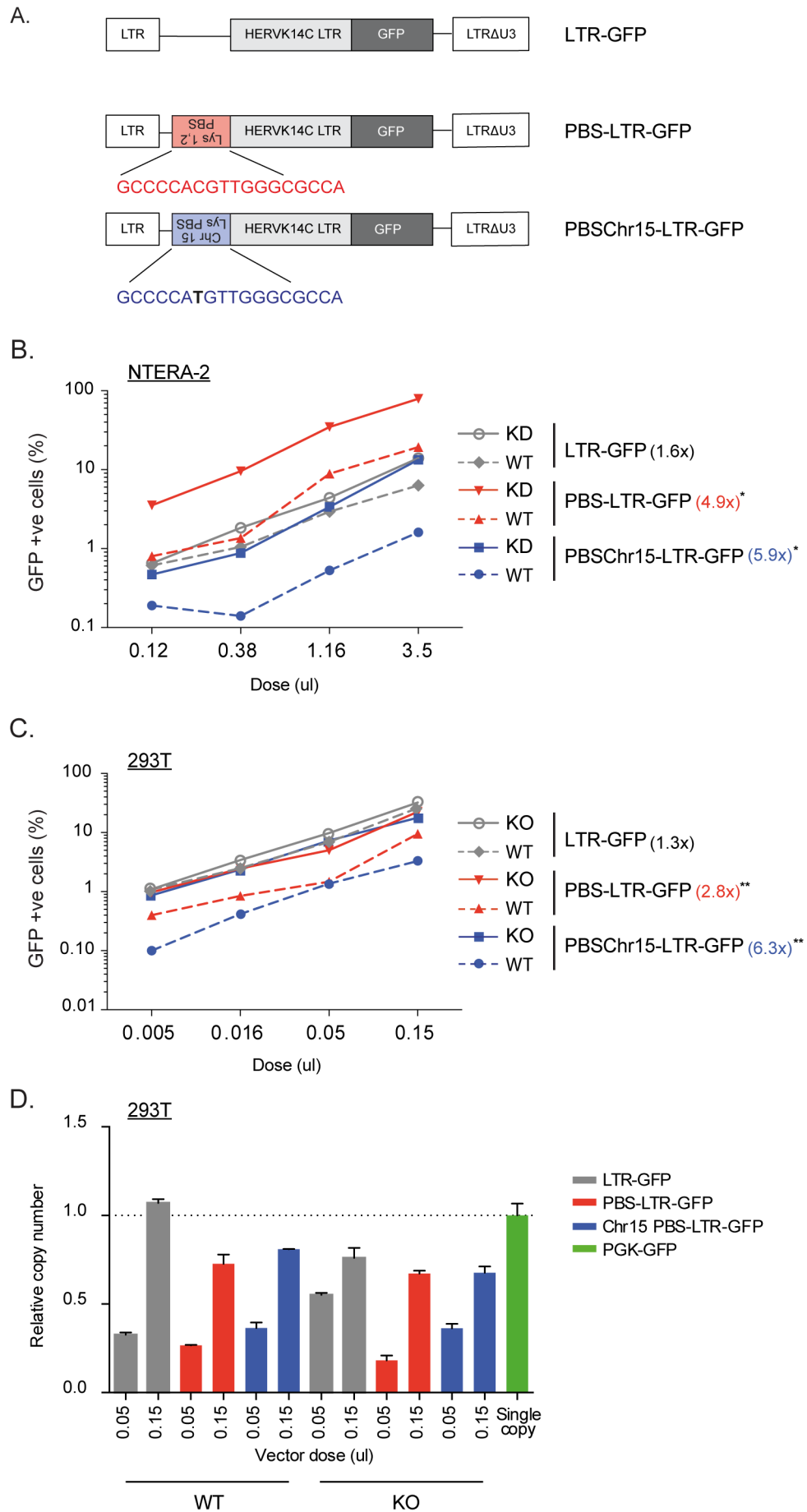


Figure 23 KAP1 is crucial for PBS-mediated *de novo* silencing of incoming ERVs.

- (A) Schematic diagram of the reporter constructs used.
- (B) NTERA-2 cells were transduced with either a KAP1 knockdown vector (KD) or a control vector (WT) and were challenged with multiple doses of the reporter vectors and GFP was assessed 72 hours post challenge. One representative experiment of 2 experiments is shown here. Data from 2 experiments was pooled for statistical tests.
- (C) WT or KAP1 KO 293T cell clones were challenged with multiple doses of the reporter vectors and GFP was assessed 72 hours post challenge. One representative experiment of 3 experiments is shown here. Data from 3 experiments was combined for statistical tests.
- (D) Taqman qPCR looking at the relative levels of integrated vector between WT and KAP1 KO 293T cells challenged with various doses of reporters. Samples were harvested at the same time as the GFP readout. A PGK-GFP transduced cell line was used as a single copy reference. Two-tailed unpaired Student T-tests were done.

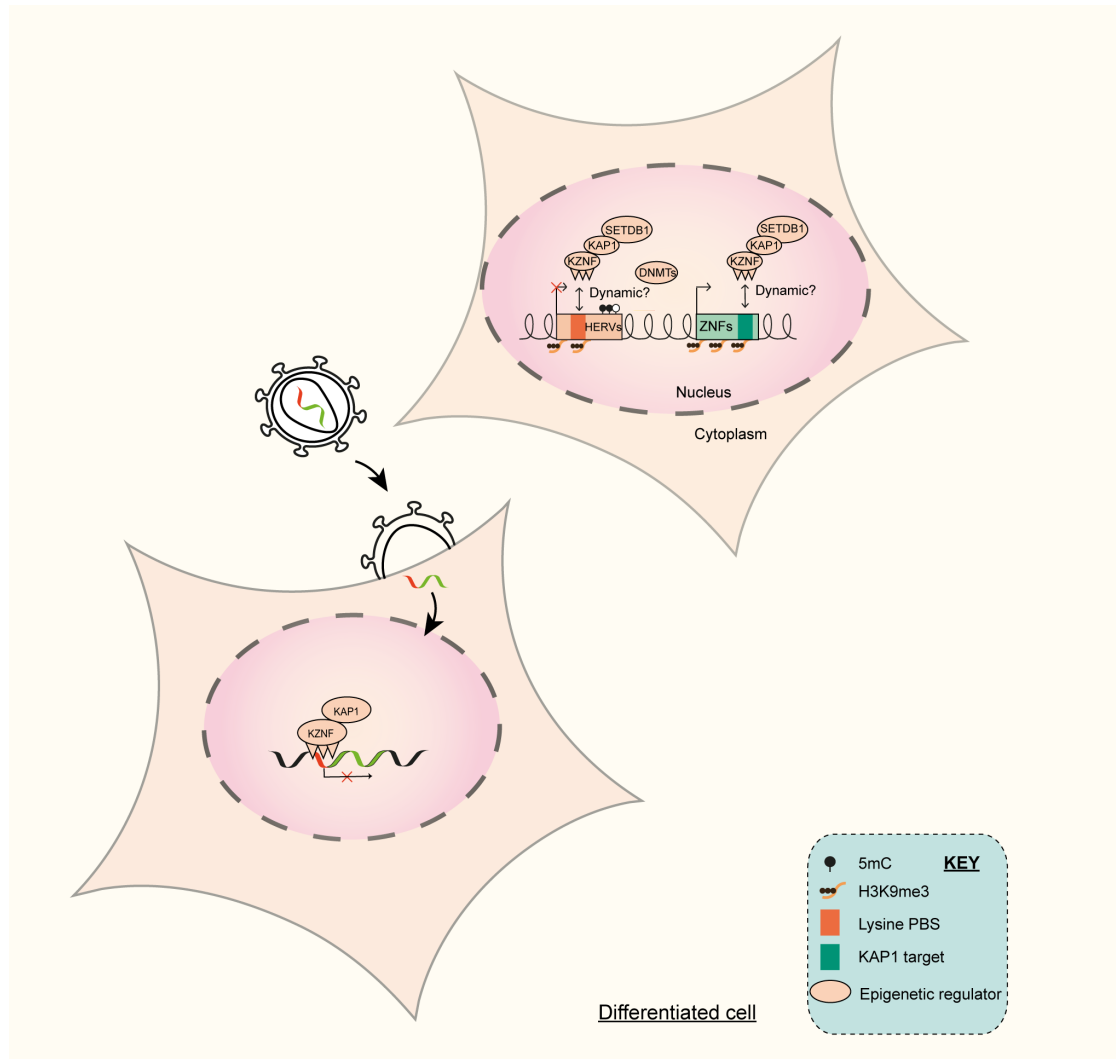


Figure 24 Model.

KAP1 represses ERVs and cellular genes (particularly ZNFs) in differentiated cells and this correlates with KAP1 binding. These loci are also enriched for SETDB1 binding and the repressive histone mark H3K9me3, which is conserved across multiple cell types. It may be that repression is not stable but dynamic as indicated, leading to potential tissue-specific activation of gene expression programmes through ERV platforms. In addition, KAP1 can restrict incoming DNA in a sequence-specific way suggesting that a subset of KZNFs are constitutively expressed in differentiated cells including the cognate KZNF for the lysine PBS. Overall, this points to the conservation and importance of the KAP1-KZNF repression pathway in differentiated cells.

Discussion I

In sum, our data illustrate that in differentiated cells, KAP1 binding and repression of a subset of ERVs is intact. The precise ERVs that become depressed upon KAP1 KO likely depend on the availability of activating transcription factors as suggested before (Wolf et al., 2015; Yang et al., 2015). KAP1 binding correlates with the presence of repressive chromatin marks that are conserved across multiple cell types. KAP1 repression can also be observed at KZNF genes suggesting that expression of ERVs and KZNFs is co-regulated. The continued expression of KZNFs in differentiated cells may allow cells to be poised to react to insults from exogenous retroviruses. Their expression may also reflect their role in fine-tuning gene expression through dynamic regulation of ERVs and ZNFs (Figure 24).

These results also show that DNA methylation perpetuated by DNMT1 is not sufficient to maintain ERVs in a silent state in adult tissues as has been proposed (Yoder et al., 1997). In contrast, we reveal an ongoing requirement for the KZNF-KAP1 pathway as suggested before from studies in mice (Ecco et al., 2016). The continued requirement of the KAP1 pathway in differentiated cells may be explained by the fact that many ERVs have been exapted by the host to mediate cell-type specific functions, for example in immune cells (Chuong et al., 2016; Collins et al., 2015; Hummel et al., 2017; Zeng et al., 2014) and potentially in the developing brain (Brattas et al., 2017; Fasching et al., 2015). KAP1-mediated regulation may allow ERVs to be expressed in a tissue-specific way or in response to environmental stimuli. Indeed, ERVs are transcribed in an organised manner during the early stages of development and, despite being categorised under the same family, individual

transcripts might have different impacts on the host (Durruthy-Durruthy et al., 2016; Goke et al., 2015; Wang et al., 2014b). The chromosomal positioning of a locus might be important in determining its fate in different developmental stages.

In comparison to HERVK14C, SVAs and LINE1s are only marginally upregulated upon KAP1 knockout indicating that dense DNA methylation may be difficult to reverse at these elements or that their activating transcription factors are lacking (Turelli et al., 2014). The enrichment of ZNF91, the cognate ZFP that recognises SVA elements (Jacobs et al., 2014), exclusively in embryonic cells points to the possibility that KAP1 and ZNF91 are not required to repress these elements in differentiated cells.

Previous studies have shown KAP1 enrichment at the 3' ends of KZNF genes on Chromosome 19 (Iyengar et al., 2011; O'Geen et al., 2007). KAP1 can also mediate the spread of HP1 β and repressive H3K9me3 histone marks at these KZNF clusters if it is artificially tethered to them (Groner et al., 2010). It has been proposed that KAP1 binding to KZNF gene bodies may primarily be to prevent recombination (Iyengar et al., 2011). However, our mRNA-seq data is consistent with a role for KAP1 in negative regulation of KZNFs, including on Chromosome 19. Interestingly, the presence of H3K9me3 on these loci does not correlate with the expression of KZNF genes upon KAP1 knockout (Kauzlaric et al., 2017). Alternatively, this observation can be explained by the loss of silent chromatin spreading from nearby ERV enhancers (or other repeats) that are KAP1 regulated as observed in recent publication in mice (Kauzlaric et al., 2017). Whether this is the case and the identities of the repeats that regulate KZNF expression in human cells are yet to be elucidated.

We found that a common feature of KAP1-regulated ERVs (HERVK14c, HERV-

S and HERV-T) was that they bound multiple transcription factors, suggesting that they function as regulatory hubs for genes in their vicinity. A recent example of ERVs regulating adjacent genes is MER41 ERVs that have been co-opted to drive the expression of ISGs (Chuong et al., 2016).

Interrogation of the expression profile of KZNFs revealed many to be widely expressed in differentiated cell types, which we also found competent for *de novo* repression of incoming DNA. Therefore, KAP1 may serve to restrict incoming human retroviruses (at least with Lys1,2, serine or threonine PBS sequences) in adult tissues, whereas in embryos it needs to silence retrotransposons more broadly, reflecting the deeper breadth of KZNFs expressed in early development. Overall, in this chapter, we have shown that:

- i) The KZNF/KAP1 pathway continues to regulate a subset of ERVs and cellular genes in human differentiated cells
- ii) This subset of ERVs are enriched for ERV1 and ERV-K elements
- iii) KAP1 can induce *de novo* repression of ERVs in a PBS-dependent manner potentially through KZNFs

Chapter 4: RKAP1 regulates ERVs in primary human cells

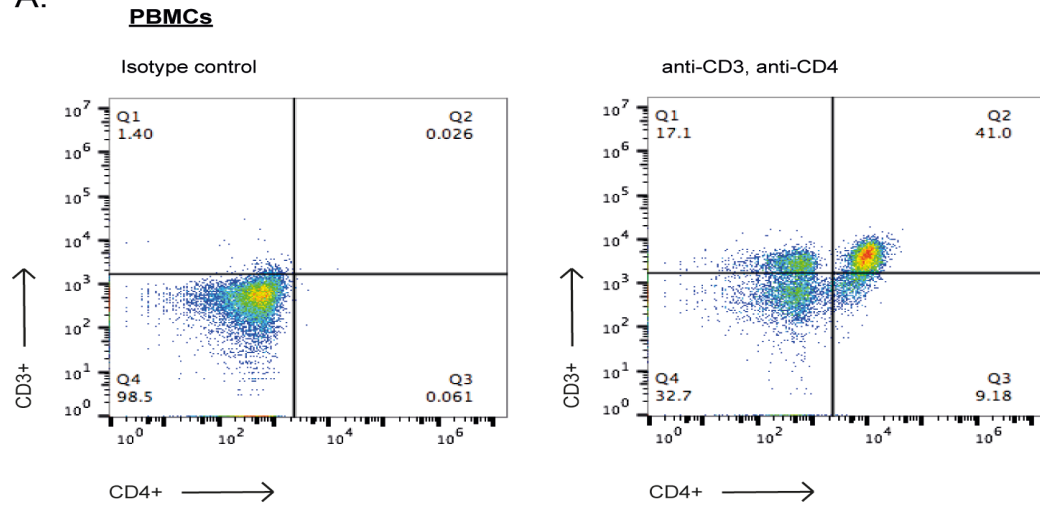
4.1 KAP1 knockdown in PBMCs leads to upregulation of retrotransposons

To extend our studies on the role of KAP1 in differentiated cells beyond cell lines, we employed primary human peripheral blood mononuclear cells (PBMCs). We isolated PBMCs either directly from fresh blood donated by healthy donors or from buffy coats as described in the methods. We found that around 60% of the population were positive for CD3, of which 70% were also positive for CD4 (Figure 25A). We then transduced the cells with shRNA-encoding lentivectors to knockdown KAP1. The knockdown efficiency was confirmed using qRT-PCR (Figure 25B and C). We observed a significant upregulation of HERVK14C and to a lesser extent LIPA4 and SVA D VNTR in both PBMC donors (Figure 25B and C).

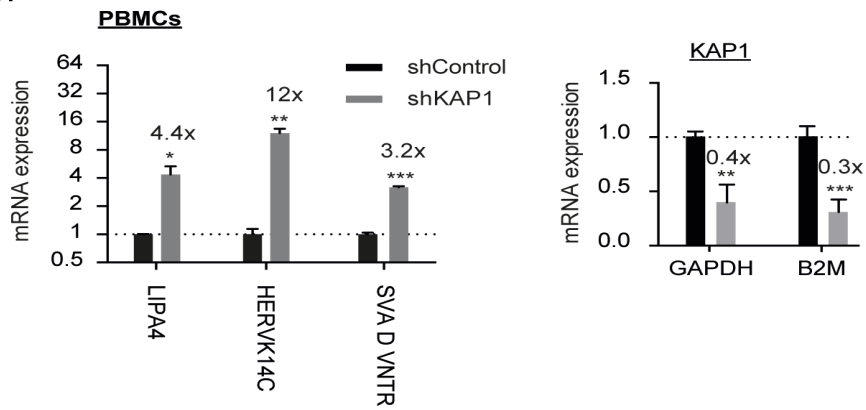
4.2 KAP1 regulates multiple KZNFs in PBMCs

Since we previously found KZNFs to be the main cellular gene targets of KAP1 repression in cell lines, we asked if this pathway was conserved in primary cells. To this end, we checked the expression level of various KZNFs in KAP1 knockdown cells and found that five out of six candidates exhibited robust and significant upregulation (Figure 25D). All of these genes are also KAP1 binding targets (Figure 17C).

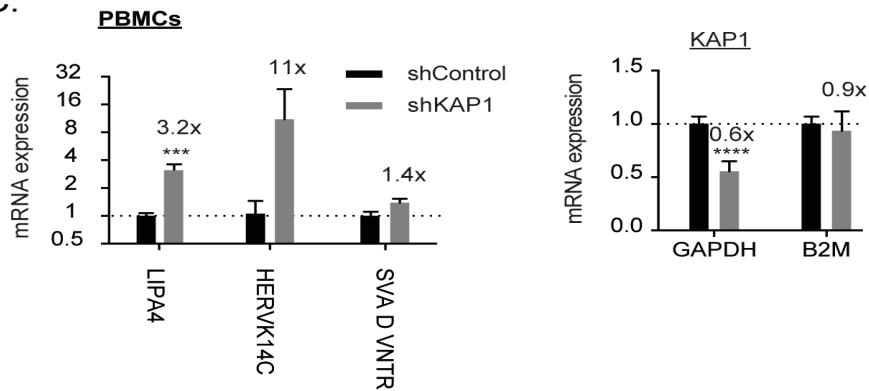
A.



B.



C.



D.

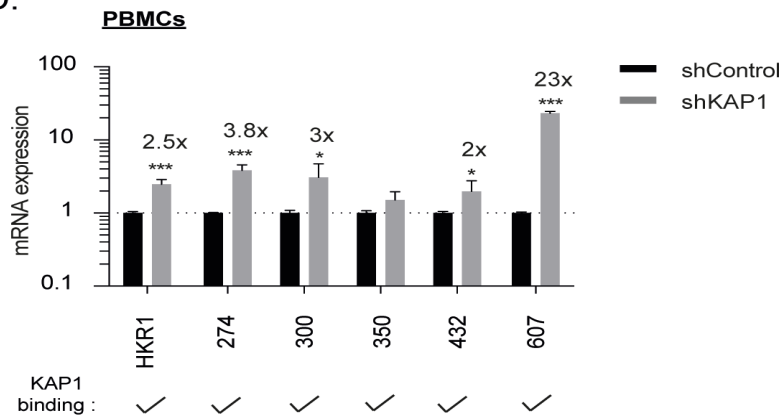


Figure 25 KAP1 repression of ERVs is conserved in PBMCs.

- (A) FACS plots showing staining of T cell markers in PBMCs.
 - (B) qRT-PCR expression of retrotransposons following KAP1 knockdown in PBMCs. Results were normalised to B2M. The level of KAP1 expression is shown in a separate graph (right) with B2M and GAPDH used as normalisers.
 - (C) The same experiment as in (B) using PBMCs from a different donor.
 - (D) The expression levels of various KZNFs in two independent experiments (B & C). B2M was used as the normaliser gene.
- Two-tailed unpaired Student T-tests were done.

4.3 KAP1 knockdown is not sufficient for robust upregulation of HERVK14C in CD4+ T cells

Since PBMCs represent a mixed population of cells, we further purified them to isolate CD4+ T cells, which formed the bulk of our PBMCs (Figure 25A), to ask if we could detect KAP1-regulation of ERVs within this population. Using our isolation procedure (see methods), we could achieve up to 95% purity (Figure 26A). We then transduced the cells with shRNA-encoding lentivectors and verified the knockdown efficiency with Western blot and qRT-PCR over several time points (Figure 26B and C). Surprisingly, despite an efficient depletion of KAP1, we did not see a robust upregulation of HERVK14C (Figure 26C). We hypothesised that DNA methylation may be enriched at HERVK14C and pose a barrier to its reactivation. In support of this possibility, we observed an enrichment of cytosine methylation at the HERVK14C LTR (Figure 26D). Finally, consistent with the lack of HERVK14C reactivation in these cells, we also saw little effect on the expression of KAP1 targeted KZNFs (Figure 26E). It's possible that KAP1 targets would have become activated over time or that the cognate KZNFs and / or an activation signal was simply missing in these cells. To partly address this, we employed our *de novo* reporter assay in these cells (see below).

4.4 KAP1 can restrict incoming ERVs through the PBS sequence in PBMCs

We took advantage of our HERVK14C reporters developed previously and transduced first PBMCs with increasing doses of the reporters. Interestingly, we

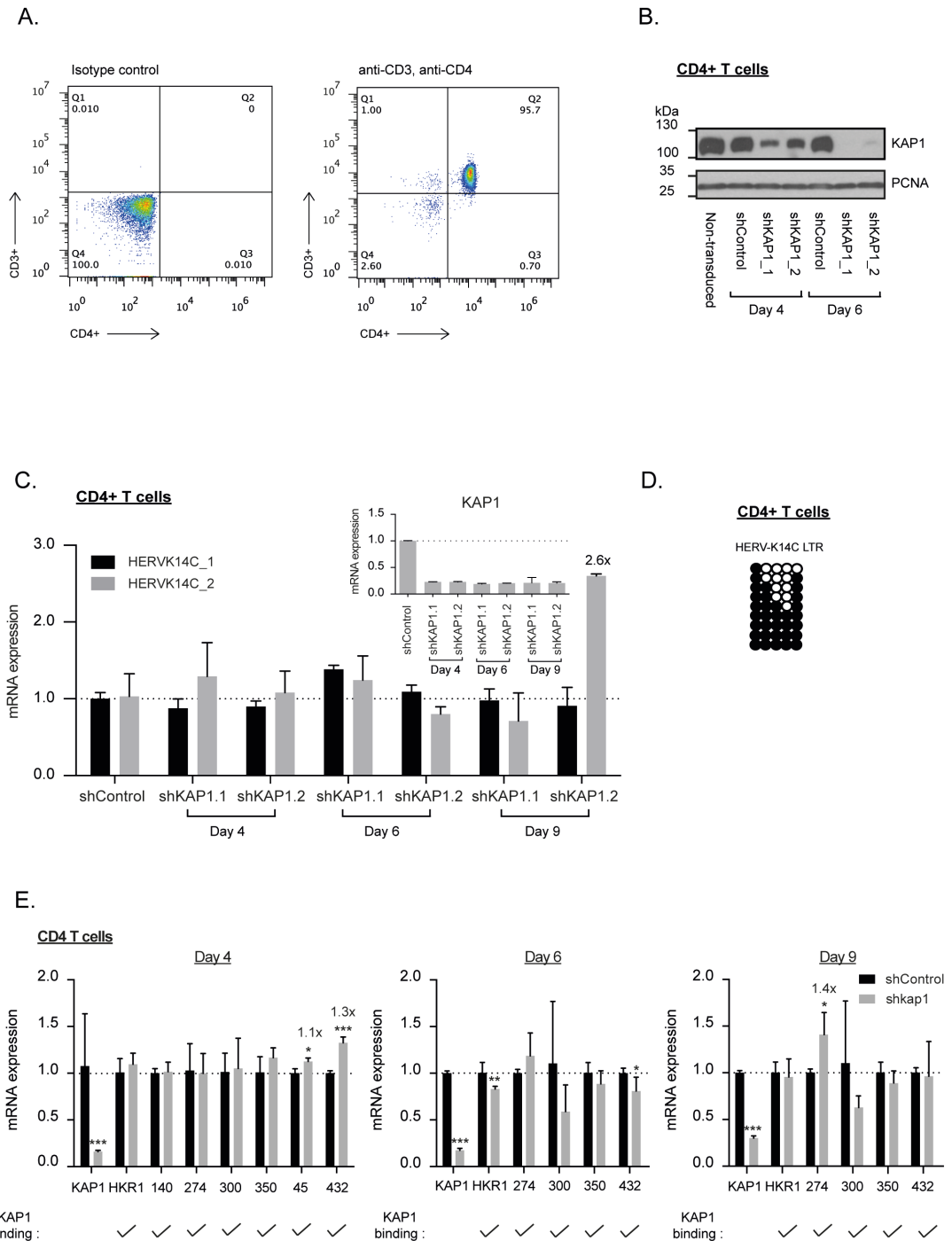


Figure 26 KAP1 depletion leads to little or no HERVK14C upregulation in CD4+ T cells.

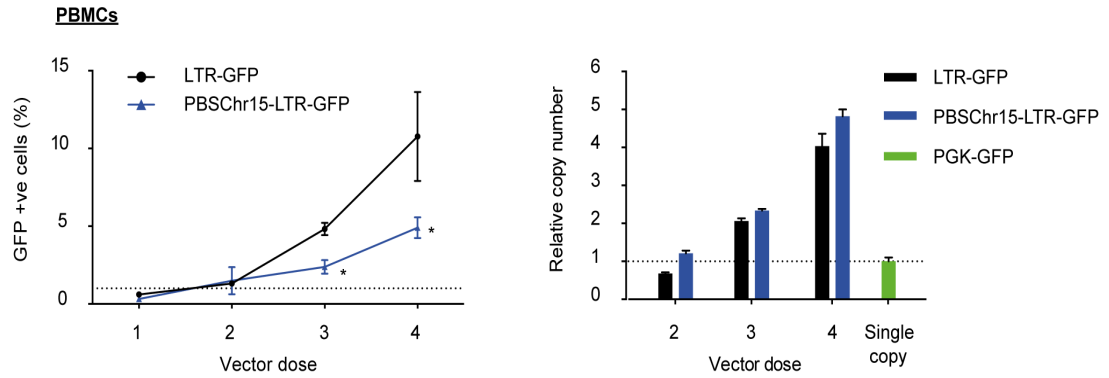
- (A) FACS plots showing the purity of CD4+ T cells used in the experiments.
- (B) Western blot showing the KAP1 knockdown efficiency in CD4+ T cells at various time-points.
- (C) qRT-PCR expression of HERVK14C following KAP1 knockdown. The inset shows the mRNA level of KAP1 post-knockdown. Results were normalised to B2M.
- (D) DNA methylation analysis of a single locus of endogenous HERVK14C LTR in CD4+ T cells.
- (E) qRT-PCR expression of various KZNF genes previously identified (Figure 17C) in KAP1 knockdown cells. Results were normalised to B2M.
- Two-tailed unpaired Student T-tests were done.

consistently saw a modest GFP repression of the PBSChr15-LTR-GFP vector compared to the control vector, LTR-GFP (Figure 27A, left). This observation was not due to a difference in vector dose as evidenced by equivalent integration of the two reporters (Figure 27A, right). In contrast, we saw no repression when we performed the same experiment in CD4⁺ T cells (Figure 27B). Intriguingly, the presence of the endogenous HERVK14C regulation phenotype appeared to mirror the presence of the reporter repression phenotype with both being present in PBMCs but absent in purified CD4⁺ T cells (Figure 25 and 26). This suggests that differences may relate to a difference in the expression profiles of KZNFs. Indeed, using the publicly available data (accession number for the data - macrophages: GSE36952; CD4⁺ T cells: GSE69549), we found that despite a large overlap in the KZNFs expressed, 118 of them are expressed in macrophages but not in CD4⁺ T cells (Figure 27C).

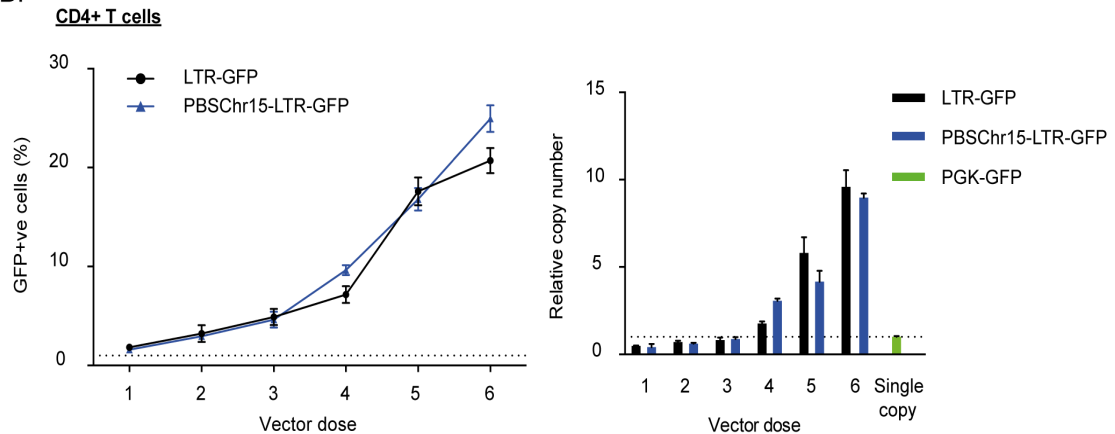
4.5 Multiple KZNFs are widely expressed in various differentiated cell types

The fact that *de novo* reporter KAP1 repression and endogenous retrotransposon and ZNF repression is intact in PBMCs suggests that a proportion of KZNFs must be expressed in all the cell types we have tested. We, therefore utilised publicly available mRNA-sequencing data from 293T cells, HeLa cells, macrophages and CD4⁺ T cells. As hypothesised, we found many KZNFs (77) to be expressed at the mRNA level across all data sets and 159 KZNFs detected in at least 2 of the cell types tested (Figure 28A). Among the 77 KZNFs, around two thirds of them are bound by KAP1 (Figure 28A). For the top 100 KZNFs expressed in 293T cells, we verified them to be expressed at the protein level using the human protein atlas and we

A.



B.



C.



Figure 27 PBS-mediated *de novo* repression of ERVs is conserved in primary cells.

- (A) PBMCs were transduced with increasing doses of GFP reporter vectors (vectors were titrated on KAP1 KO 293T cells prior to use in primary cell experiments) and GFP was assessed 72 hours post transduction (left). The bar chart (right) shows vector integration measured by Taqman qPCR with GFP primers. A 293T cell line with a single vector copy integrant (PGK-GFP) was used as a control (right). A representative of 3 experiments is shown here. Data from 3 experiments was pooled for statistics.
- (B) The same experiment as in (A) but with CD4+ T cells. A representative of 2 experiments is shown here. Data from 2 experiments was pooled for statistics.
- (C) Venn diagram showing the KZNF expression profile between macrophages and CD4+ T cells. Two-tailed unpaired Student T-tests.

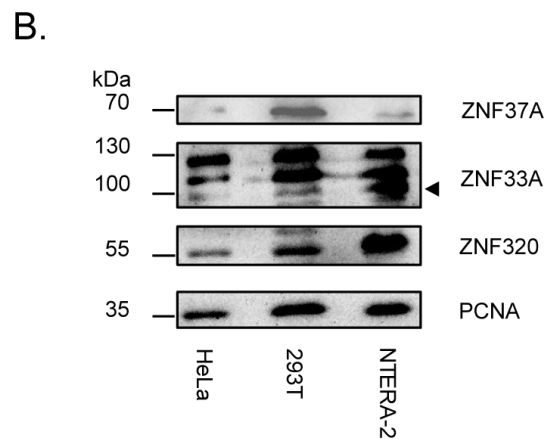
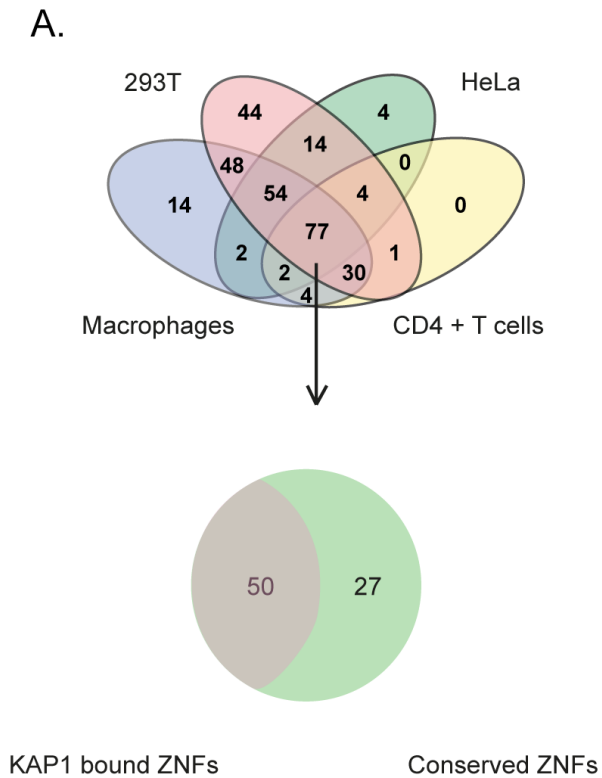


Figure 28 ZNFs are expressed widely at the mRNA level.

(A) Venn diagram showing the unique and overlapping ZNFs expressed in multiple cell types (top). Venn diagram showing the proportion of conserved ZNFs that are KAP1 bound (50 out of 77) (bottom).

(B) Western blots confirming the expression of three KZNFs in both undifferentiated (NTERA-2) and differentiated cells (293Ts and HeLa cells).

additionally verified several of them to be expressed in our cell types of interest using Western blot (Figure 28B).

Discussion II

In conclusion, our data show that KAP1 is required to maintain ERVs repressed and to regulate ZNFs in primary human cells (Figure 29, top). Hundreds of KZNFs are widely expressed in various differentiated cell types, illustrating the widespread importance of the KAP1-KZNF pathway in regulating ERVs. However, differences exist between cell types in sequence-specific regulation, since we were unable to detect KAP1 repression in CD4⁺ T cells (Figure 29, bottom). This likely relates to subtle differences in the repertoire of ZNFs expressed and potentially also to differences in activating transcription factors. It may also be the case that KAP1 repression is dynamic as indicated (Figure 29, top) and in activated CD4⁺ T cells, transcription of assessed KAP1 targets is “ON” despite detectable cytosine methylation consistent with previous observation (Ecco et al., 2016).

The human immune system is an interactive network, the cellular components of which are transcriptionally dynamic, allowing them to adapt to infections (Parkin and Cohen, 2001). Such plasticity suggests that epigenetic regulation of these components likely involves histone modifications rather than DNA methylation. It has previously been shown that ERVs play a role in immune cells (Chuong et al., 2016; Collins et al., 2015; Hummel et al., 2017; Zeng et al., 2014) as well as KAP1 (Collins et al., 2015; Hummel et al., 2017). From our own PBMC data, KAP1 knockdown resulted in the upregulation of HERVK14C and KZNFs. Interestingly, in contrast to cell lines, LIPA4 and SVA elements were also upregulated. This might

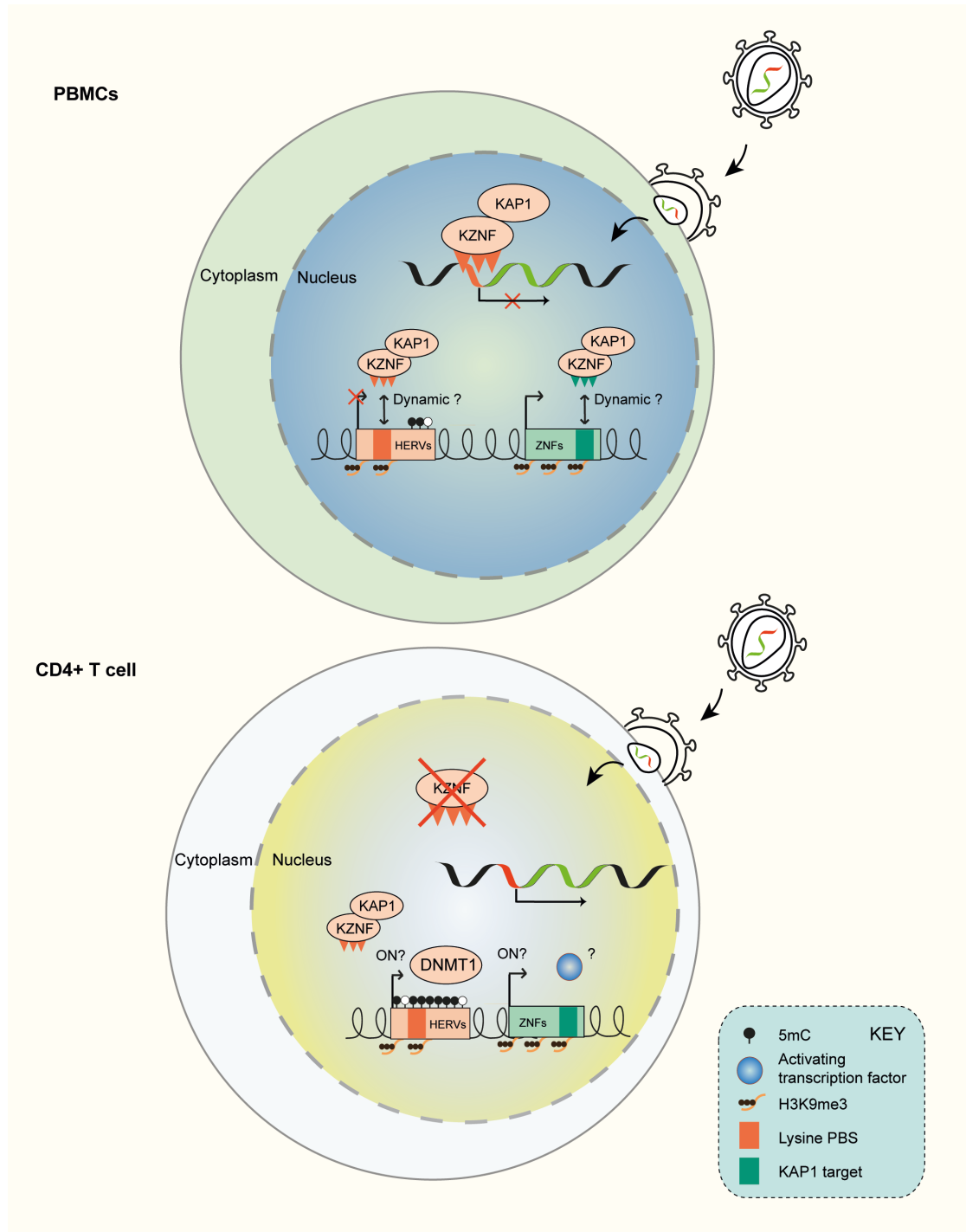


Figure 29 Model.

Like in differentiated cell lines, KAP1 is required to repress ERVs and to regulate ZNFs in primary human cells, a process that may be dynamic. KAP1 *de novo* repression of foreign DNA through the PBS sequence is also intact in PBMCs suggesting that the components of this pathway are expressed at a sufficient level to restrict incoming exogenous retroviruses. CD4+ T cells, in contrast, lack lysine PBS repression. This may be due to low / absent expression of the cognate KZNF(s), despite the fact that we detect 118 KZNFs to be expressed in these cells. We also did not observe reactivation of ERVs and ZNFs in KAP1-depleted CD4+ T cells. This may be due to knockdown efficiency or redundant repression pathways or suggest that ERV and ZNF transcription is highly active at baseline.

reflect differences in DNA methylation between cancer lines and PBMCs as it is well-documented that the landscape of DNA methylation is disrupted in cancer cells (Kulis and Esteller, 2010). The fact that the DNA methylation inhibitor 5-Aza affects SVAs more in cell lines than in PBMCs (see Chapter 3) indicates that SVAs retain DNA methylation in cell lines but may not be so in PBMCs.

The fact that we detect KAP1-regulation in PBMCs but not in CD4+ T cells suggests that the co-culture of immune cells is important for the upregulation phenotype and may influence the profile of transcriptional factors expressed (Goke et al., 2015; Grow et al., 2015; Lu et al., 2014b)

Of note, KAP1-repression is only modest in differentiated cells, which may relate to the low expression levels of KZNFs. Intriguingly, the mouse KZFP, Zfp809 is rapidly degraded at the protein level in differentiated cells, restricting its potency mainly to embryonic cells (Lee and Bieniasz, 2007; Wang and Goff, 2017). An open question that remains is whether KZNFs could become induced in response to viral infections, interferon treatment or other stimuli to make them more potent in differentiated cells thus supporting their potential function as restriction factors. Overall, in this chapter, we have shown that:

- i) KZNFs/KAP1 are involved in the repression of ERVs in PBMCs and in the initiation of *de novo* silencing of incoming ERVs
- ii) Many KZNFs are universally transcribed and translated in multiple cell types, suggesting they may function broadly to regulate ERVs

Chapter 5: Epigenetic control of ERVs prevents viral mimicry

5.1 DNA methylation and histone acetylation regulate retrotransposons

Recent publications (Chiappinelli et al., 2015; Roulois et al., 2015) have proposed a potential working mechanism for the cancer drug, 5-Aza, involving DNA demethylation of ERVs and innate immune activation, which then contributes to cancer cell death along with adaptive anti-tumour immunity (Figure 30A). These publications support a dominant role for DNA methylation in the suppression of retrotransposons in adult tissues, in line with previous articles (Rowe et al., 2013a; Rowe et al., 2010; Walsh et al., 1998). DNA methylation is triggered by the KZNF-KAP1 pathway (Rowe et al., 2013a), as well as through KAP1-independent mechanisms (Leung et al., 2011). While it is known that aberrant activation of ERVs can induce viral mimicry and innate immune activation (Chiappinelli et al., 2015; Roulois et al., 2015), almost nothing is known about the epigenetic factors required to maintain repression and DNA methylation of immune-triggering ERVs. In this chapter, we asked if KAP1 and related epigenetic factors were necessary to prevent viral mimicry in differentiated human cells (Figure 30A). We first assessed if KAP1-regulated ERVs were DNA methylated since immune-triggering ERVs are known to be regulated through DNA methylation (Chiappinelli et al., 2015; Roulois et al., 2015). To test this, we treated HeLa cells with either the DNMT inhibitor, 5-Aza, or the HDAC inhibitor, TSA for comparison (Figure 30B-D). Following 5-Aza treatment, KAP1-regulated HERVK14C and SVA D VNTR were upregulated as measured by qRT-PCR (Figure 30C). SVA D VNTR was more potently affected, as expected, since these

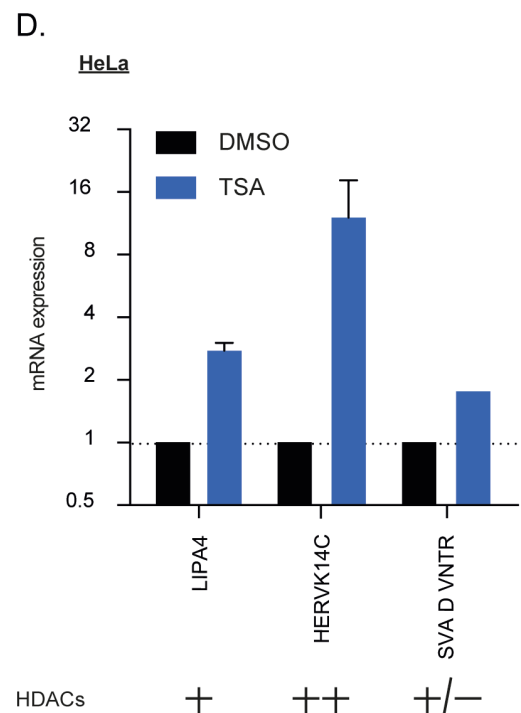
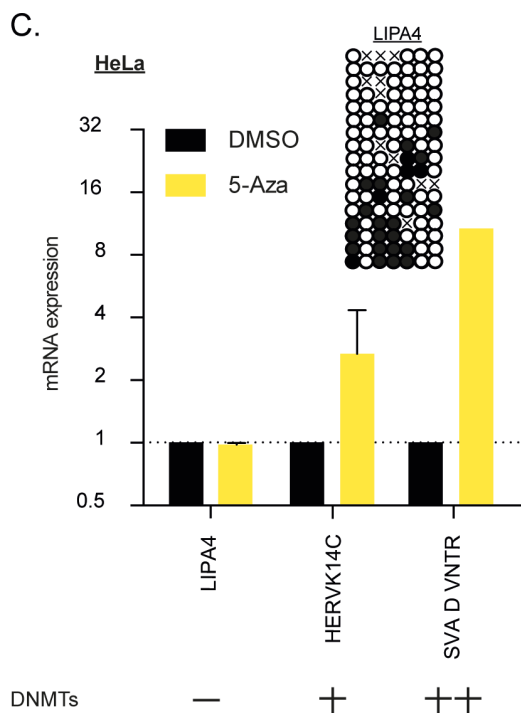
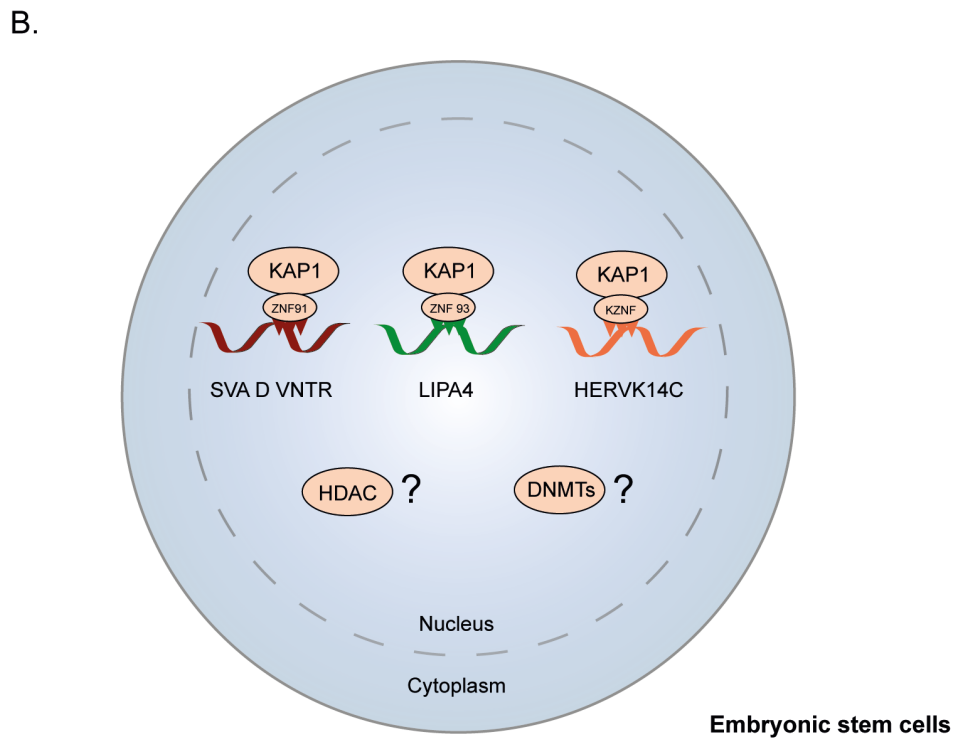
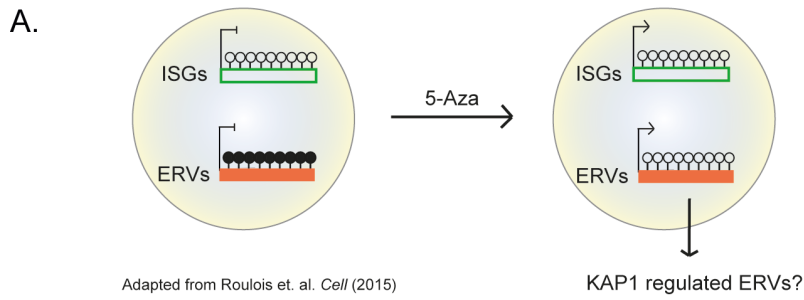


Figure 30 KAP1-regulated retrotransposons are subject to DNA methylation.

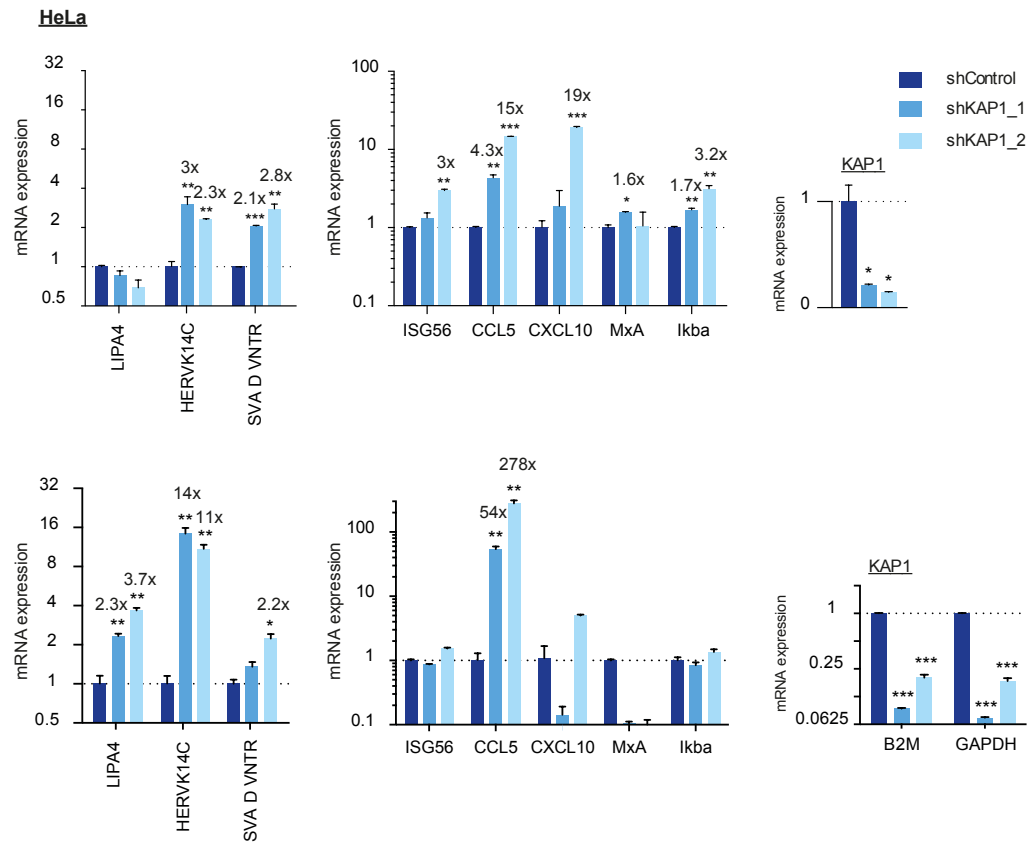
- (A) Model proposed to explain how 5-Aza treatment leads to induction of ISGs. A filled lollipop represents a methylated CpG while an open lollipop represents an unmethylated CpG.
- (B) Model illustrating the question we are asking here of whether KAP1-regulated retrotransposons are DNA methylated and sensitive to 5-Aza.
- (C) Expression of KAP1-regulated retrotransposons was measured by qRT-PCR following 5-Aza drug treatment of HeLa cells. B2M was used as the normalising gene. - : dependency on DNA methylation not detected; + : dependency on DNA methylation. The inlay shows the DNA methylation status of LIPA4 elements in 293T cells. 'x' denotes the missing CpG at the specified location of the LIPA4 promoters.
- (D) The same experiment as in (C) but here with TSA as the drug. The dependency on HDACs for repression is scored below. -/+ : inconclusive.

elements are enriched for DNA methylation (Figure 16E) (Turelli et al., 2014). L1PA4 was unaffected, in contrast, perhaps reflecting its low DNA methylation status (Figure 30C, see inlay). TSA treatment, which inhibits class I and II HDACs, also induced upregulation of these retrotransposons, particularly HERVK14C (Figure 30D), showing that KAP1-regulated ERVs are subject to histone deacetylation as well as DNA methylation, as predicted from previous work (Ivanov et al., 2007; Rowe et al., 2010).

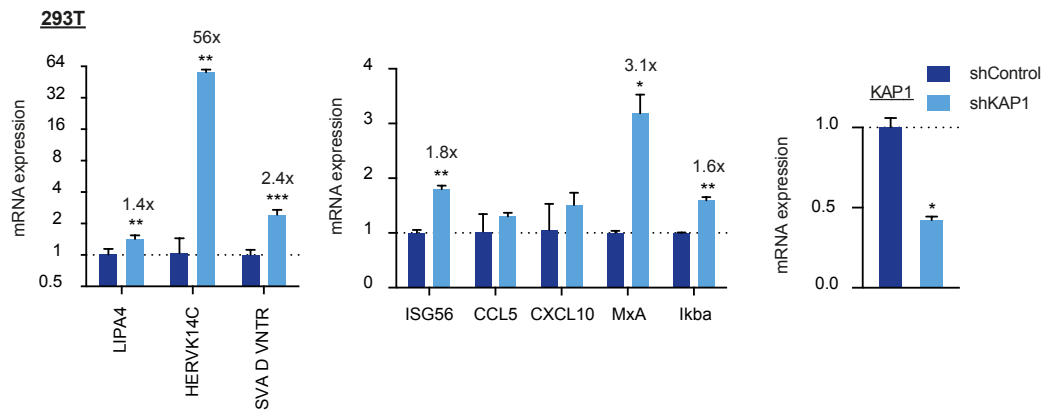
5.2 KAP1 contributes to the prevention of an aberrant innate immune response

We then asked if KAP1-depletion would lead to aberrant immune activation as well as induction of retrotransposon expression. Indeed, we found that loss of KAP1 in HeLa cells led to the overexpression of HERVK14C and SVA D VNTR elements as noted previously (see Chapter 3) and was sufficient to trigger the activation of several chemokines, mainly CCL5 (Figure 31A). Results were not reproducible in 293T cells and PBMCs (Figure 31B and C), however, despite consistent retrotransposon induction (see Figure 25B and C for PBMC retrotransposon expression data), suggesting that KAP1 depletion alone is not sufficient for global induction of interferon-stimulated genes (ISGs). It is likely that the phenotype is dependent on a range of factors including the timing, the knockdown efficiency and / or the type and magnitude of retrotransposons reactivated.

A.



B.



C.

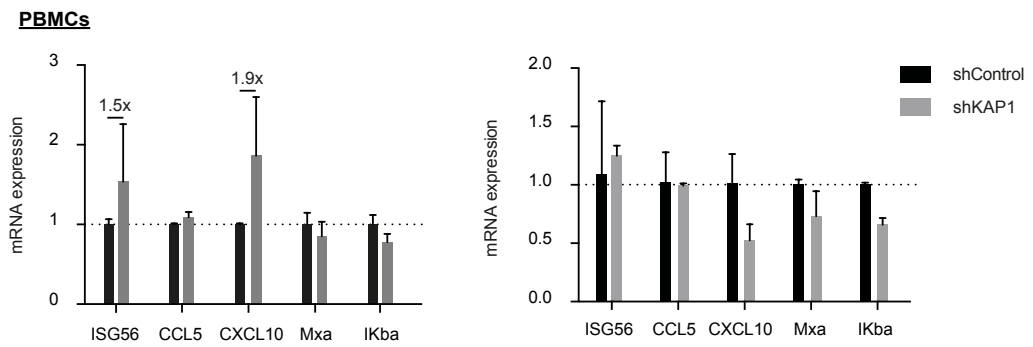


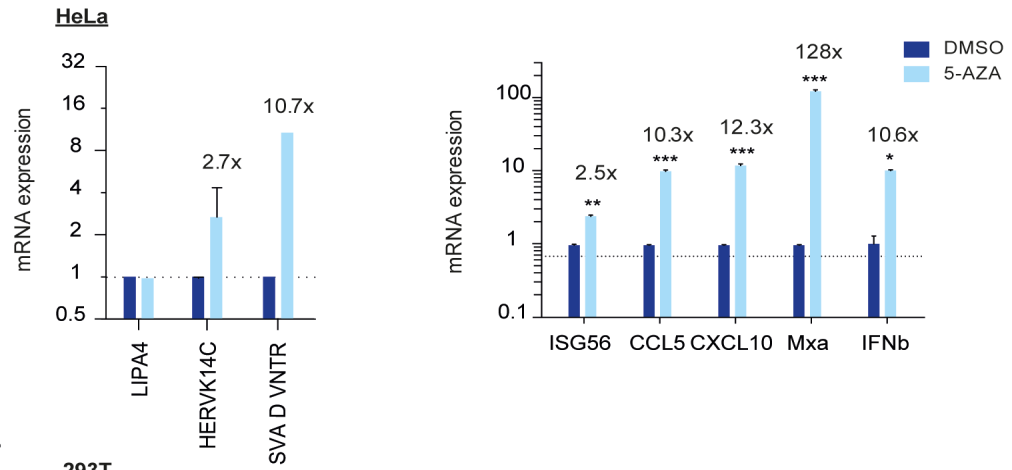
Figure 31 KAP1 depletion affects expression of ISGs.

- (A) Expression was measured by qRT-PCR of KAP1-regulated retrotransposons (left), ISGs (middle), and KAP1 (right) following KAP1 knockdown in HeLa cells. Results were normalised to B2M and / or GAPDH. Two sets of independent experiments are shown here (top and bottom).
- (B) The same as in (A) but here with 293T cells. One representative experiment of 3 experiments is shown here.
- (C) PBMCs from Figure 3.1B and C were assessed for ISG induction using qRT-PCR. Results were normalised to GAPDH.
Two-tailed unpaired Student T-tests were done.

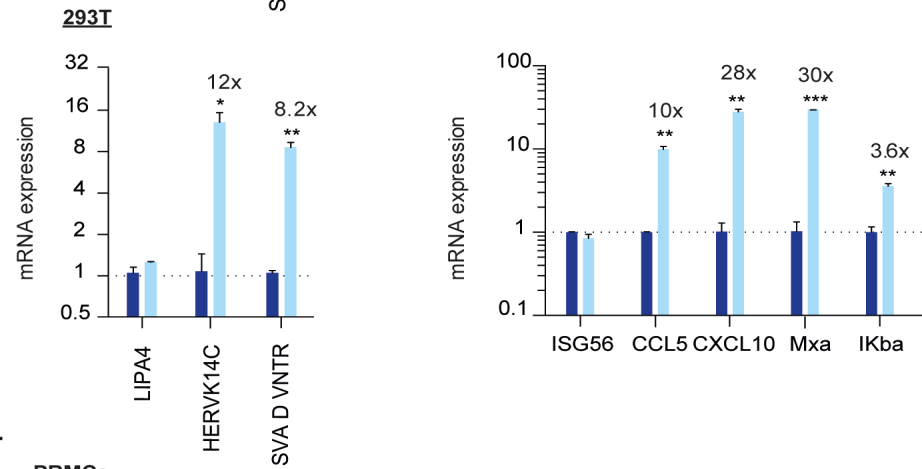
5.3 5-Aza treatment triggers an ISG response

We next asked if 5-Aza treatment would lead to aberrant immune activation as has been proposed to be the case in cancer cell lines (Chiappinelli et al., 2015; Roulois et al., 2015). We therefore treated HeLa and 293T with 5-Aza for 96 hours and checked the relative expression levels of repeats and ISGs. As previously observed (Figure 30C), SVA D VNTR and HERVK14C were upregulated in both cell types (Figure 32A and B) and this was accompanied by a clear induction of ISGs (Figure 32A and B). It is unknown whether 5-Aza treatment, which is used in cancer patients stimulates an innate immune response only in cancer cells or indiscriminately in all dividing cells. We therefore treated PBMCs with the drug and surprisingly, we could also observe ISG induction that was significant albeit lower (Figure 32C). Importantly, this was accompanied by an overexpression of KAP1-regulated retrotransposons, suggesting a link between KAP1, DNA methylation and the prevention of viral mimicry in differentiated cells. Of note, however, the induction of viral mimicry may result from loss of H3K9me3 (that can also be KAP1-dependent (Rowe et al., 2010)), rather than loss of DNA methylation since 5-Aza has been reported to also lead to H3K9me3 depletion (Komashko and Farnham, 2010). We tested the drug TSA for comparison and we found that this drug had only a modest effect on the expression of ISGs (Figure 32D). One difference was that SVA D VNTR elements were not activated, suggesting that these retrotransposons may be one of the elements that produce immuno-stimulatory nucleic acids.

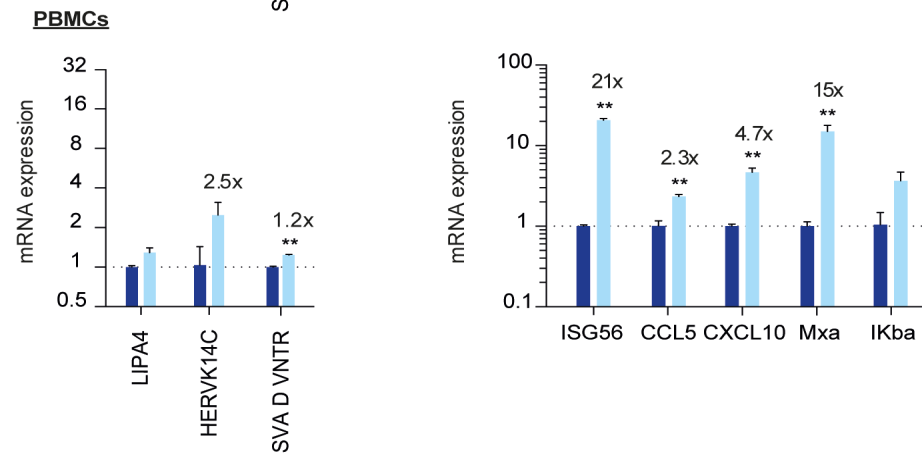
A.



B.



C.



D.

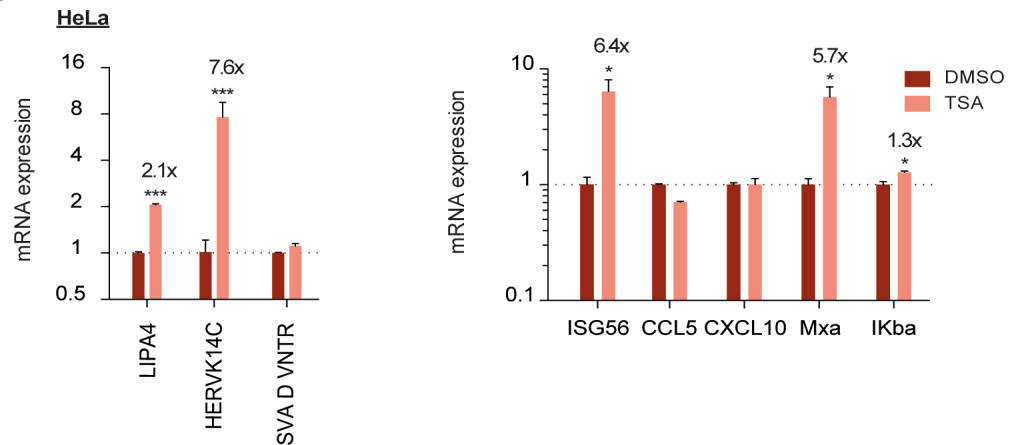


Figure 32 5-Aza treatment leads to an upregulation of KAP1-regulated retrotransposons and robust ISG induction in cancer cell lines and primary cells.

- (A) Expression as measured by qRT-PCR of KAP1-regulated retrotransposons (left) and ISGs (right) following 5-Aza treatment of HeLa cells. Results were normalised to B2M and / or GAPDH. One representative experiment of at least 2 independent experiments is shown here.
- (B) The same experiment as in (A) but here in 293T cells.
- (C) The same experiment as in (A) but here in PBMCs.
- (D) The same experiment as in (A) but here with TSA treatment.
- Two-tailed unpaired Student T-tests were done.

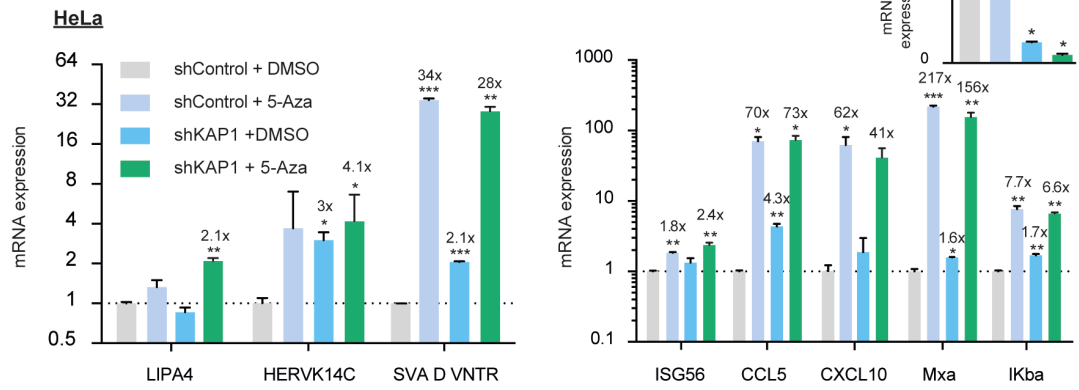
5.4 There is redundancy between KAP1 repression and DNA methylation

We asked if KAP1 depletion and 5-Aza treatment would have an additive effect on the induction of retrotransposons and ISGs. To address this, we knocked down KAP1 and combined this with 5-Aza drug treatment in both cell lines followed by qRT-PCR. Knockdown efficiencies were first verified using qRT-PCR (Figure 33A and B, see inlays). As before, the phenotype was striking with 5-Aza, whereas KAP1 depletion had only a modest effect on retrotransposons and ISGs and combined treatments had no additive nor synergistic effect (Figure 33A and B). This suggests that DNA methylation is the major gatekeeper that prevents viral mimicry through retrotransposons and KAP1 likely acts in the same pathway but is largely redundant with DNA methylation. For comparison, we performed the same experiments but with the drug TSA. KAP1 knockdown or TSA treatment led to modest upregulation of HERVK14C and the combination showed only a marginal additive effect on HERVK14C expression (Figure 33C and D) with no additive effect on ISG induction (Figure 33D). This further supports the idea that viral mimicry results from retrotransposons that are epigenetically silent through DNA methylation rather than histone deacetylation alone.

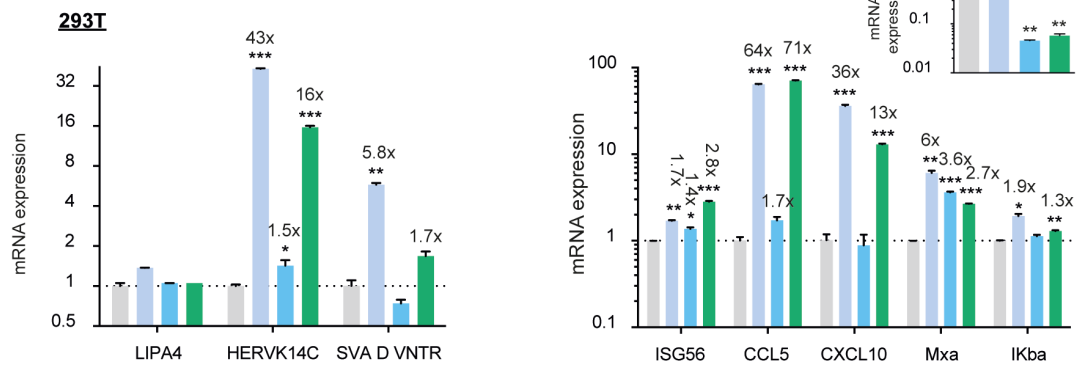
5.5 KAP1 is involved in the maintenance of DNA methylation at ERVs

Here, we have shown evidence that DNA methylation is key to the prevention of viral mimicry (Figure 32 and 33). KAP1 is linked to DNA methylation because *de novo* KAP1 repression leads to *de novo* DNA methylation of ERVs in embryonic cells (Rowe et al., 2013a). We asked here if KAP1 is necessary to maintain DNA methylation at ERVs. A limitation of our KAP1 knockdown experiments is that

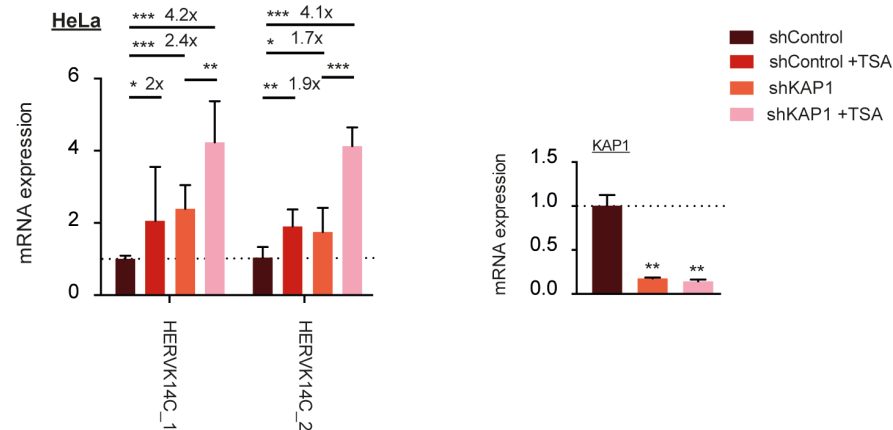
A.



B.



C.



D.

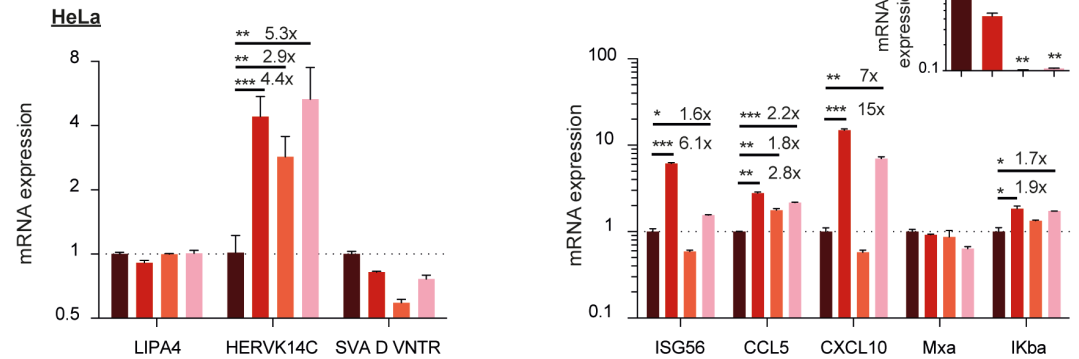


Figure 33 Combining KAP1 knockdown with 5-Aza treatment did not have an additive effect.

- (A) Expression as measured by qRT-PCR of KAP1-regulated retrotransposons (left) and ISGs (right) following KAP1 knockdown and /or 5-Aza treatment of HeLa cells. The inlay shows the knockdown efficiency of KAP1. Results were normalised to GAPDH. One representative experiment of 2 is shown here.
- (B) The same experiment as in (A) but here in 293T cells. Results were normalised to B2M or GAPDH. A representative of 2 experiments is shown here.
- (C) and (D) The same experiment as in (A) but here TSA treatment is used instead of 5-Aza and ISGs were only measured in (D). Results were normalised to B2M or GAPDH. One representative experiment of 3 experiments (C) or 2 experiments (D) is shown.
- Two-tailed unpaired Student T-tests were done.

KAP1 may still be expressed at sufficient levels to maintain DNA methylation at ERVs. We therefore employed our KAP1 knockout HeLa cell clones to assess DNA methylation at endogenous HERVK14C. Indeed, we found that DNA methylation was reduced at HERVK14C in knockout compared to wild type cells (Figure 34).

5.6 The role of epigenetic regulators that KAP1 collaborates with in the prevention of viral mimicry

In addition to KAP1, we also explored the role of other epigenetic modifiers in protecting cells from aberrant innate immune activation. We focused on factors that KAP1 collaborates with, namely the histone methyltransferase SETDB1 and the recently discovered HUSH complex (Tchasovnikarova et al., 2015). SETDB1 mediates H3K9me3 and partners with KAP1 to maintain H3K9me3 at retrotransposons (Matsui et al., 2010; Rowe et al., 2013b). The HUSH complex is composed of three components: TASOR, encoded by *FAM208A*, MPP8, encoded by *MPHOSPH8* and periphilin encoded by *PPHLN1*, and interacts with SETDB1 and MORC2 (Tchasovnikarova et al., 2017). The HUSH complex is implicated in H3K9me3 spreading (Tchasovnikarova et al., 2017; Tchasovnikarova et al., 2015) and our unpublished work shows that it collaborates with KAP1 to repress retrotransposons including HERVK14C in embryonic cells (Figure 35A and B). Interestingly, in embryonic NTERA-2 cells, HERVK14C repression appears to be dependent on SETDB1, TASOR and periphilin but not on MPP8 (Figure 35B). However, we have been able to document a requirement for MPP8 in repression of LINE1 elements in both human and mouse embryonic cells (unpublished) so all three HUSH components are implicated in retrotransposon repression. We therefore depleted all

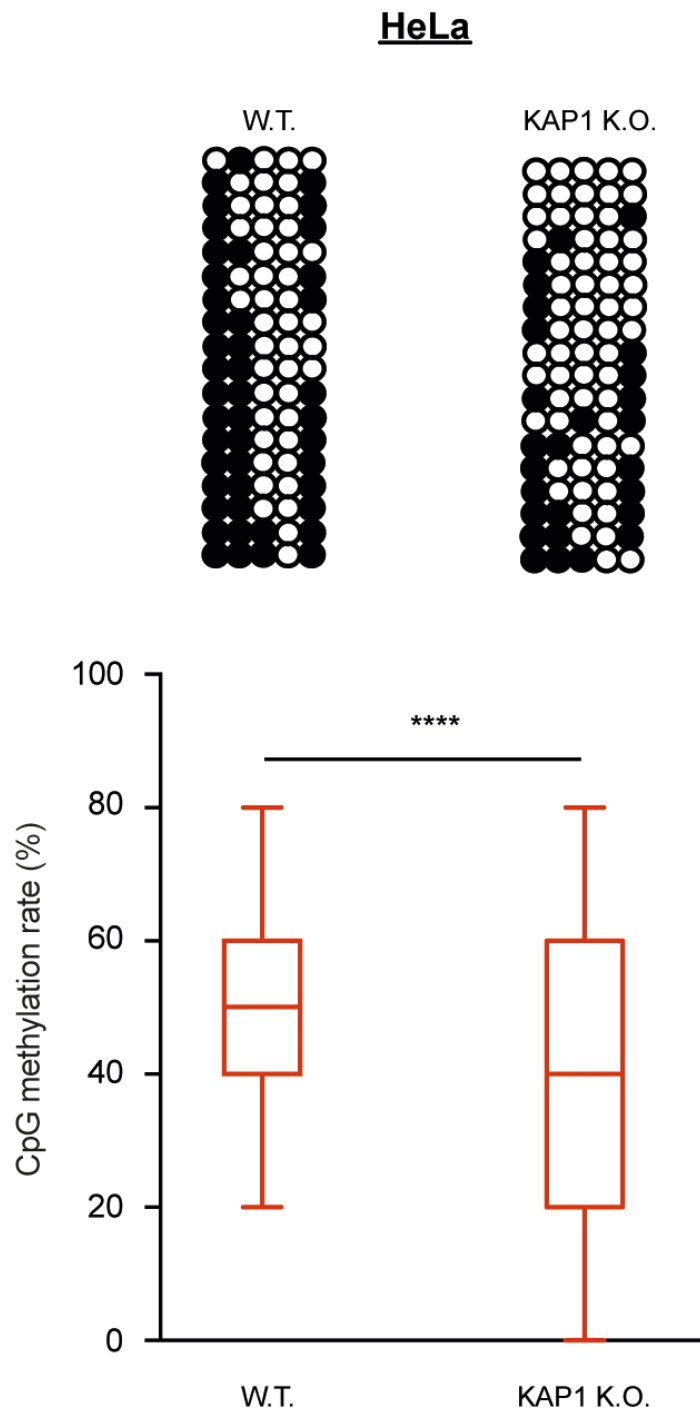
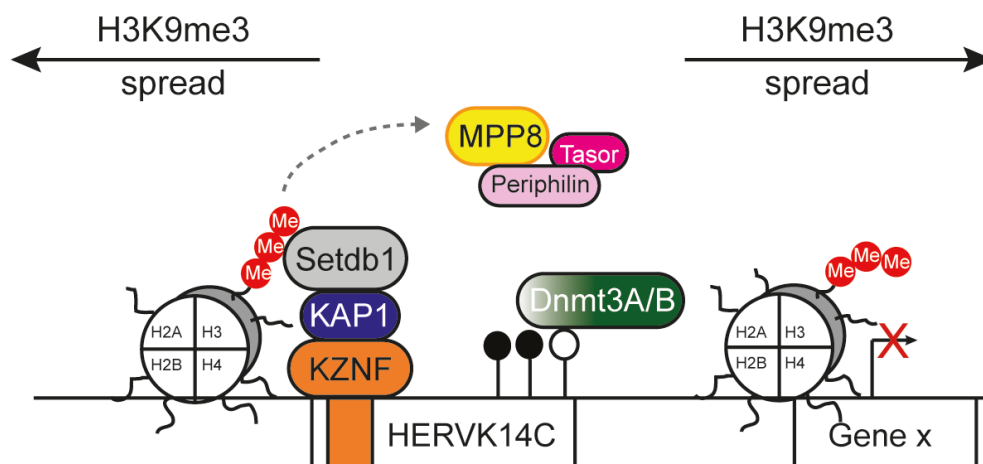


Figure 34 KAP1 is required to maintain DNA methylation at HERVK14C.

KAP1 knockout leads to a reduction in DNA methylation at the HERVK14C LTR on Chromosome 15. The box plots show the quantification of DNA methylation over all sequences and they represent the minimum value, first quartile value, median, third quartile value and maximum value from the bottom whisker to the top whisker respectively. Two-tailed unpaired Student T-tests were done.

A.



B.

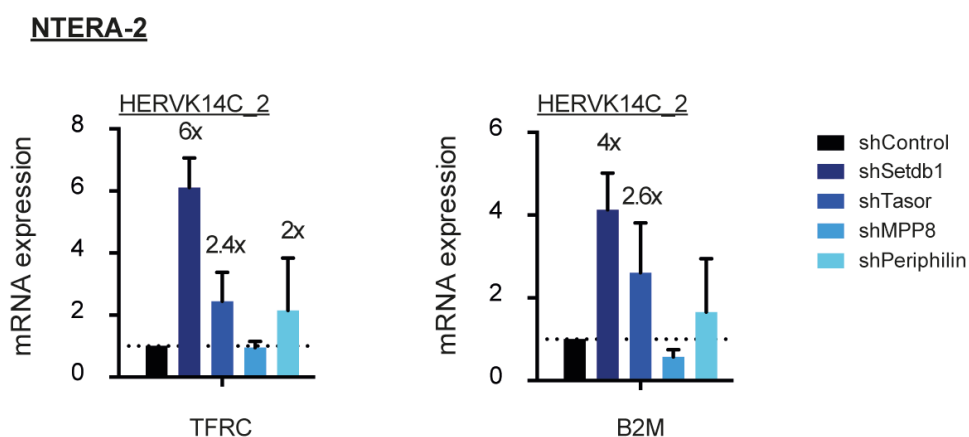


Figure 35 Multiple epigenetic modifiers are involved in repressing ERVs.

(A) Working model of KAP1 repression made from unpublished work, which shows that KAP1 partners up with the HUSH complex as well as with SETDB1 to maintain heterochromatin repression of retrotransposons, including HERVK14C.

(B) Expression of HERVK14C as measured by qRT-PCR following SETDB1 and HUSH component knockdown in NTERA-2 cells. Results were normalised to TFRC or B2M.

the above components in 293T cells and measured expression of retrotransposons and ISGs. In the case of KAP1, we used two hairpins, one targeting the UTR (labelled UTR) and one targeting the CDS (labelled shKAP1) (Figure 36A and B).

Results (Figure 36A-E) showed that KAP1 and to a lesser extent SETDB1 and periphilin depletions led to a significant overexpression of retrotransposons. Likewise, KAP1 also led to a small but significant induction of ISGs and to a lesser extent periphilin and TASOR. Of interest, CCL5 was upregulated in both TASOR and periphilin knockdowns (Figure 36D and E). MPP8 depletion, in contrast, led to a dramatic induction of ISGs and this was interestingly accompanied by a parallel significant downregulation of retrotransposon expression (Figure 36F). This phenotype of ISG induction and retrotransposon downregulation following MPP8 depletion was confirmed in at least two other independent experiments. These results suggest that MPP8 plays a unique and integral role in preventing viral mimicry or that the striking phenotype observed with MPP8 depletion relates to the knockdown efficiency, which is around tenfold greater than the others. The fact that retrotransposons are downregulated may reflect a global anti-viral state in these cells following massive ISG induction. Alternatively, retrotransposons may not be directly responsible for ISG induction here, as we do not know the identity of the PAMP.

5.7 MPP8 protects cells from aberrant ISG activation

We next assessed if the magnitude of the MPP8 phenotype related to the knockdown efficiency. We therefore transduced cells with different doses of knockdown vector. We confirmed the knockdown efficiency via qRT-PCR and

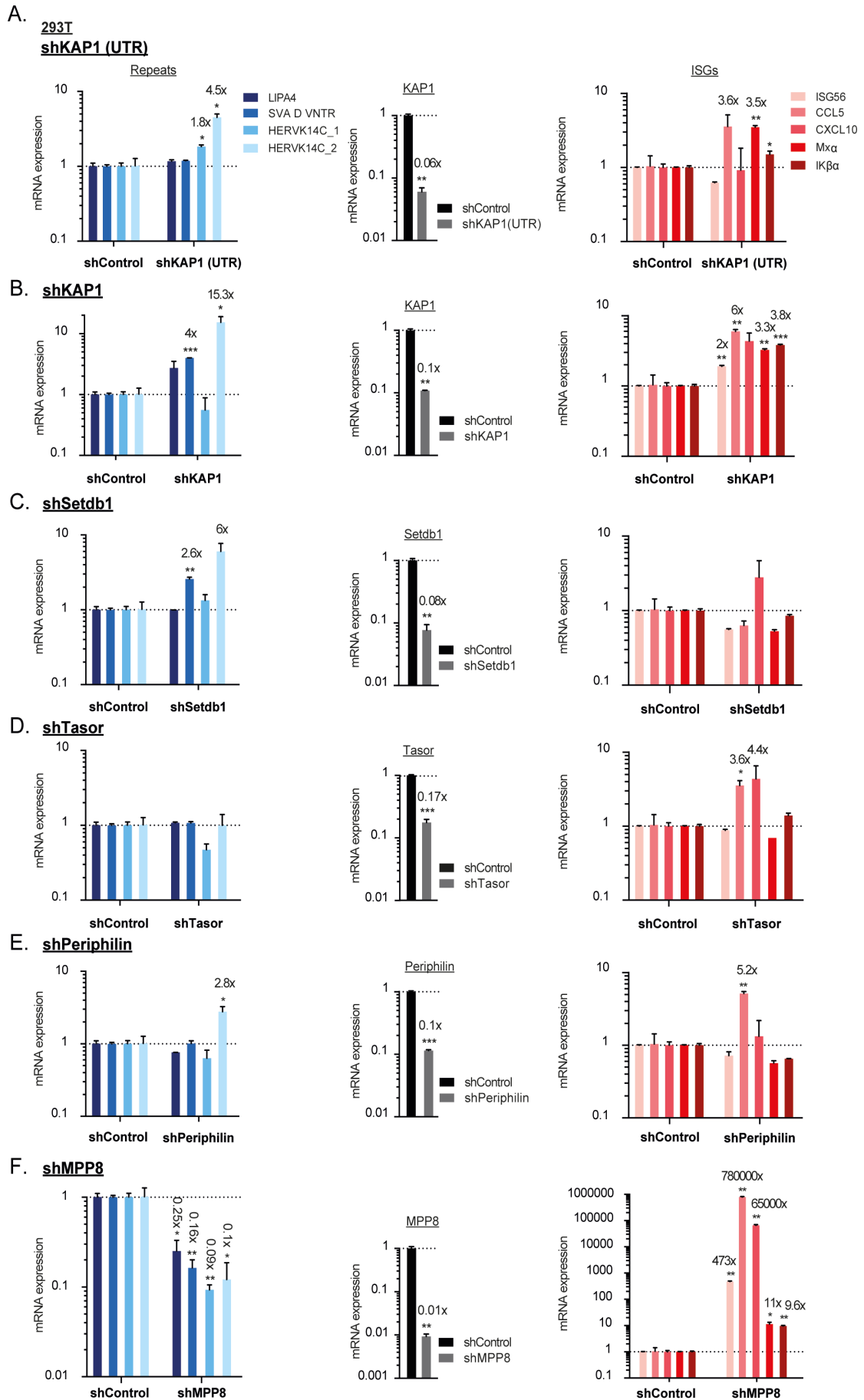


Figure 36 MPP8 prevents viral mimicry and innate immune activation.

(A – F) Expression as measured by qRT-PCR of KAP1-regulated retrotransposons (left), the stated mRNAs depleted (middle) as well as ISGs (right) following knockdown of stated epigenetic modifiers. Results were normalised to GAPDH or B2M. One representative experiment of at least 2 independent experiments is shown here. Two-tailed unpaired Student T-tests were done.

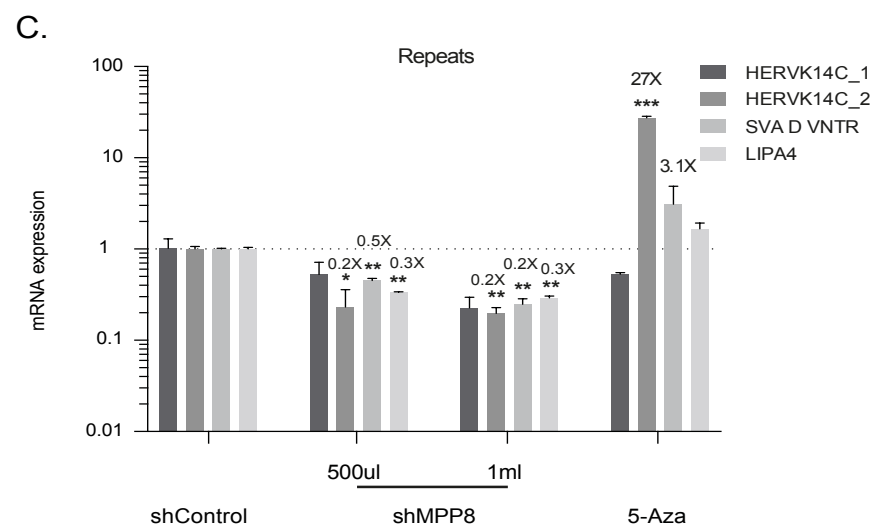
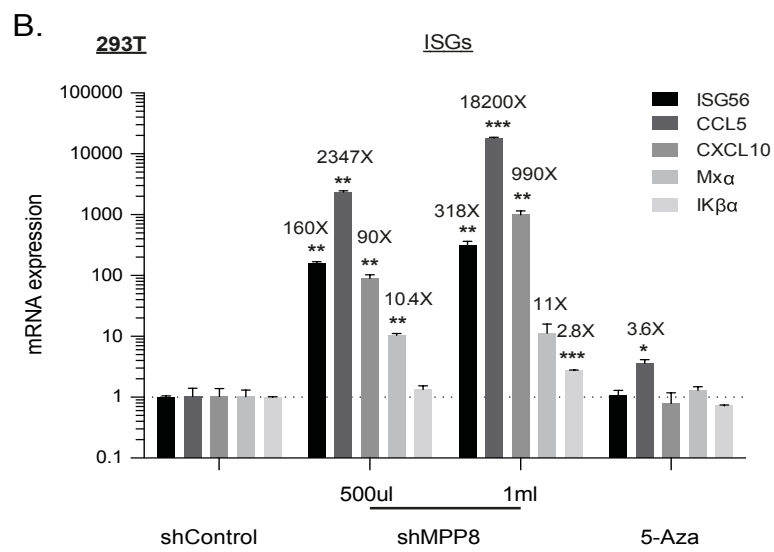
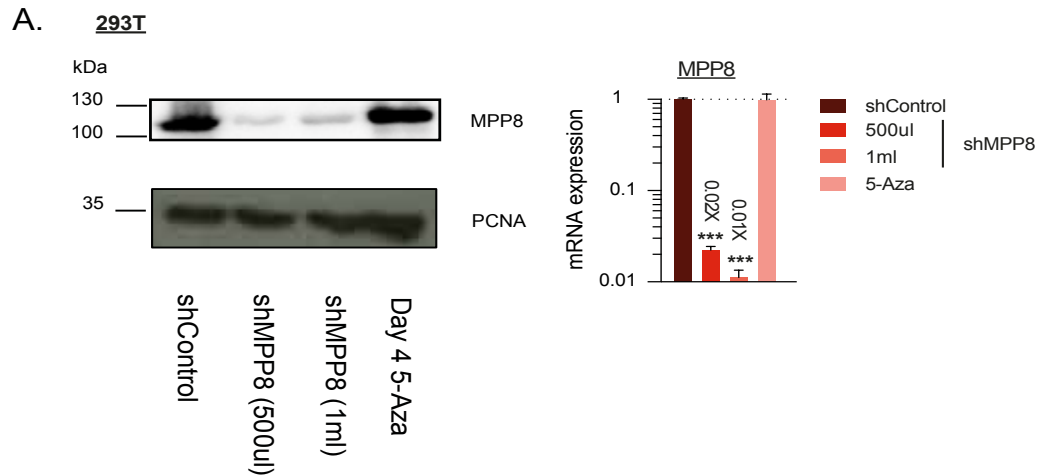


Figure 37 MPP8 depletion leads to a striking induction of ISGs while KAP1-regulated retrotransposons are downregulated.

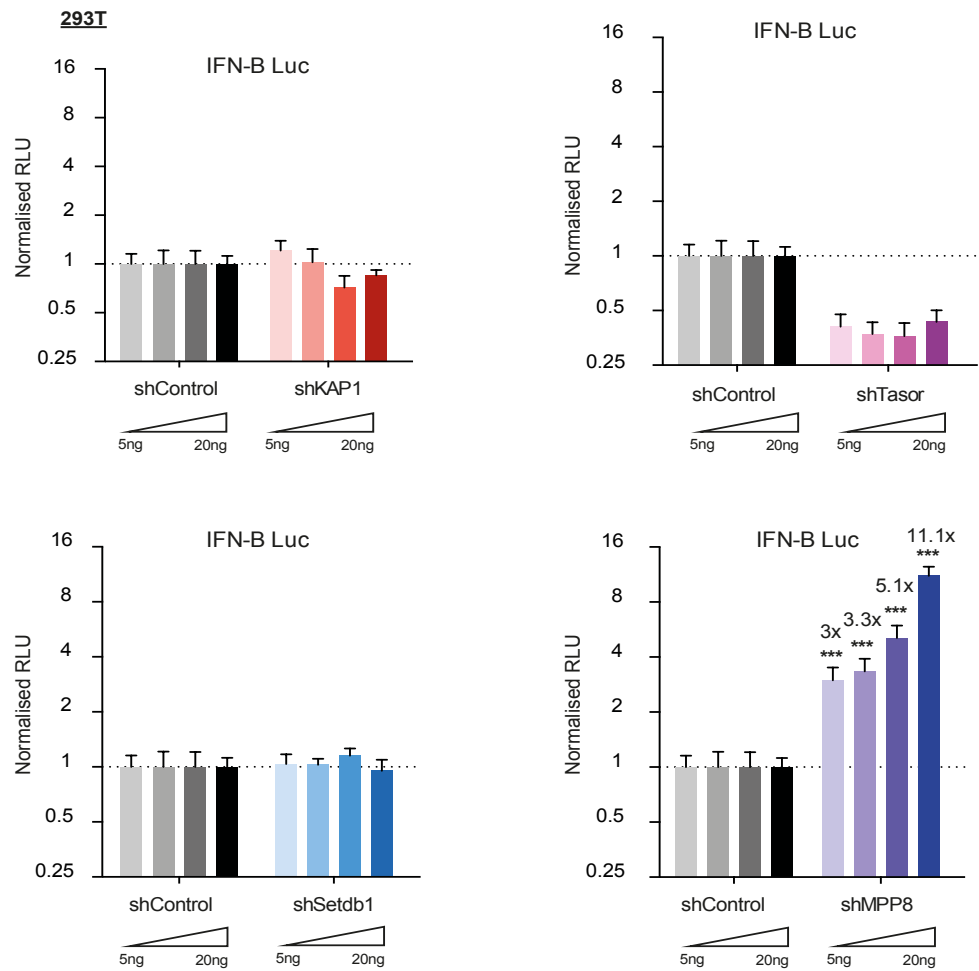
- (A) Western blot and qRT-PCR showing MPP8 knockdown efficiency. Results were normalised to B2M.
- (B) Expression of ISGs as measured by qRT-PCR following MPP8 depletion. 5-Aza was used as a positive control. Results were normalised to GAPDH. One representative experiment of 2 experiments is shown here.
- (C) Expression of KAP1-regulated retrotransposons as measured by qRT-PCR. Results were normalised to B2M.
- Two-tailed unpaired Student T-tests were done.

Western blot (Figure 37A). MPP8 depletion was extremely robust at both vector dilutions and we measured it to be 50 fold and 100 fold at the 500ul and 1ml amounts of knockdown vector, respectively (Figure 37A). As before (Figure 36), we observed a striking upregulation of ISGs upon MPP8 knockdown along with a downregulation of retrotransposons (Figure 37B and C). Importantly, the phenotype of ISG induction did relate to the knockdown efficiency because the higher MPP8 depletion correlated with more pronounced ISG upregulation (Figure 37B). We next used a luciferase based reporter that is driven by an IFN- β promoter to reinforce our findings. We, therefore transfected the knockdown cells that were depleted for the stated epigenetic modifiers with increasing doses of the luciferase reporter and measured luciferase 2 days later. We found that the reporter was only activated in the MPP8 knockdown cells (Figure 38A) and in cells loaded with Sendai virus, which was used as a positive control (Figure 38B). The fact that there was no effect on the reporter in KAP1-depleted cells despite there being modest activation of some ISGs in these cells as detected by qRT-PCR (Figure 36) may relate to the different sensitivities of the assays and / or differences in genes measured. These data support an integral role for MPP8 in preventing an innate immune response and show that the magnitude of the response is tightly linked to the knockdown efficiency.

5.8 MPP8 depletion activates classical type I IFN signalling involving IRF3 and JAK-STAT

To reinforce the link between MPP8 and the protection from immune reactivity, we sought to demonstrate that MPP8 knockdown activates classical type I IFN

A.



B.

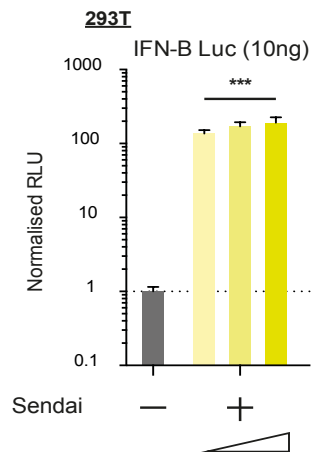


Figure 38 MPP8 depletion leads to activation of an IFN β reporter.

(A) 293T cells transduced with different knockdown vectors were transfected with increasing amounts of a luciferase-expressing plasmid driven by an IFN β promoter. The luciferase signal (RLU) was normalised to the renilla signal.

(B) Wild type 293T cells were transfected with 10ng of the reporter plasmid used in (A) in the presence of increasing doses of Sendai virus. The luciferase signal (RLU) was normalised to the renilla signal.

Two-tailed unpaired Student T-tests were done.

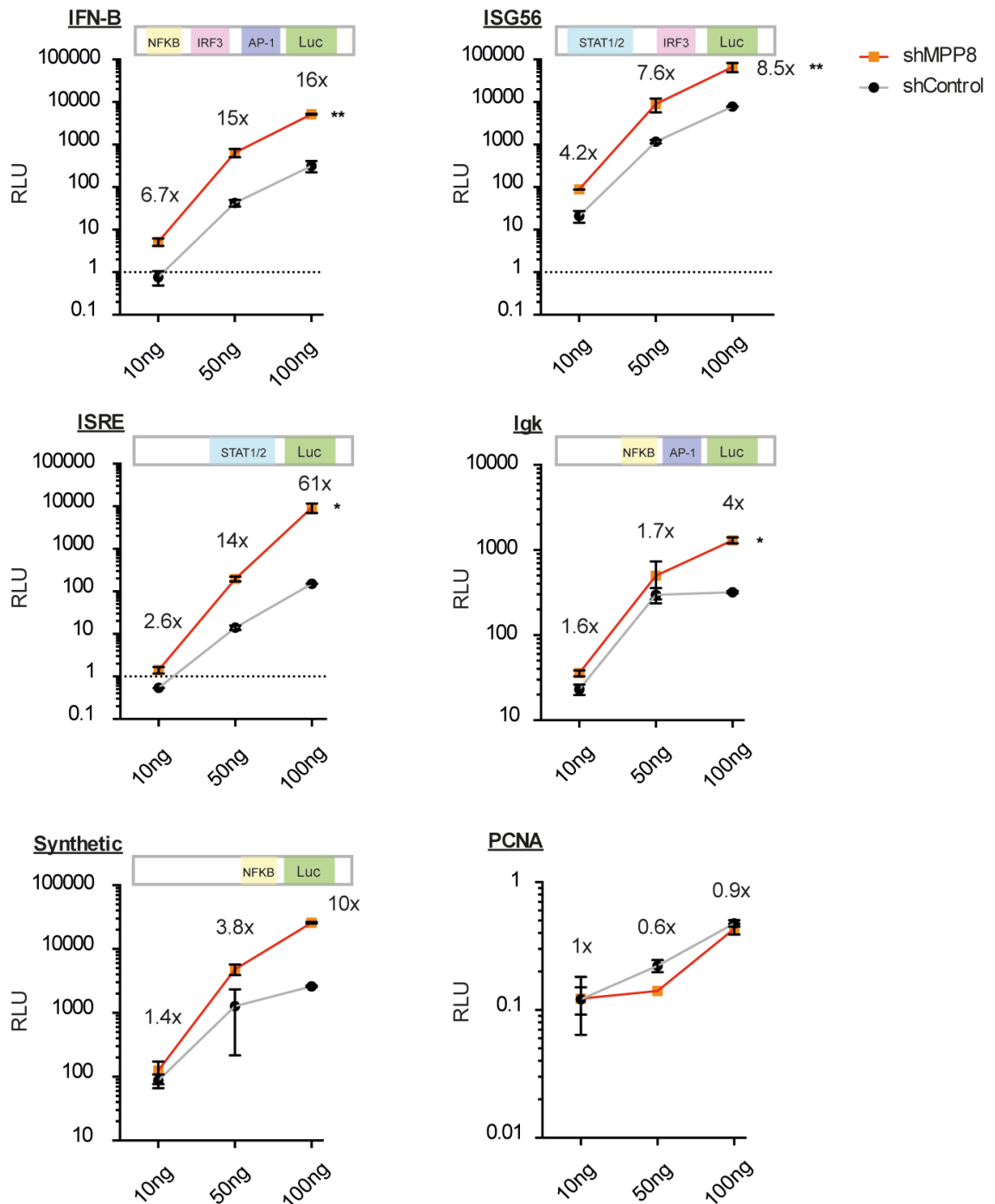


Figure 39 MPP8 depletion leads to activation of IRF3 and JAK-STAT dependent reporters.

The same experiment as in Figure 4.8 but here various luciferase reporters were used as stated. A map of each reporter is shown at the top of the graphs. The PCNA control does not encode luciferase and serves as a control for background fluorescence. Two-tailed unpaired Student T-tests were done.

signalling. We therefore transfected MPP8-depleted cells with a panel of luciferase plasmids driven by various promoters. We observed significant and consistent activation of IFN- β , ISG56 and ISRE promoters while the IgK and synthetic promoters were less affected because with the latter two promoters, a difference between MPP8 knockdown and control cells was only observed at one dose of plasmid (Figure 39). According to the transcription factor binding sites of these reporters (Figure 39), these results indicate that IRF3 and JAK-STAT signalling are active in MPP8-knockdown cells, both of which are upstream of ISG induction.

Discussion III

Overall, our data suggest an active role of epigenetic modifiers in protecting cells from an unwanted immune response through sensing of endogenously-derived PAMPs and activation of type I IFN signalling (Figure 40). We provide evidence here to show that MPP8 is an integral gatekeeper preventing immune reactivity, presumably due to its function in heterochromatin spreading at retrotransposons. KAP1 plays a more minor role in this immune protection perhaps reflecting the redundancy of epigenetic pathways (Figure 40) or simply the lower knockdown efficiency of KAP1 compared to MPP8. Our experiments with the 5-Aza drug show that DNA methylation is crucial to prevent an innate immune response and we have shown that KAP1 is involved in the maintenance of this mark at least at some retrotransposons. This is an emerging field with many questions unanswered. Future work will focus on key questions such as what are the precise retrotransposons that can mimic viruses once active, whether they can simulate RNA or DNA sensing or

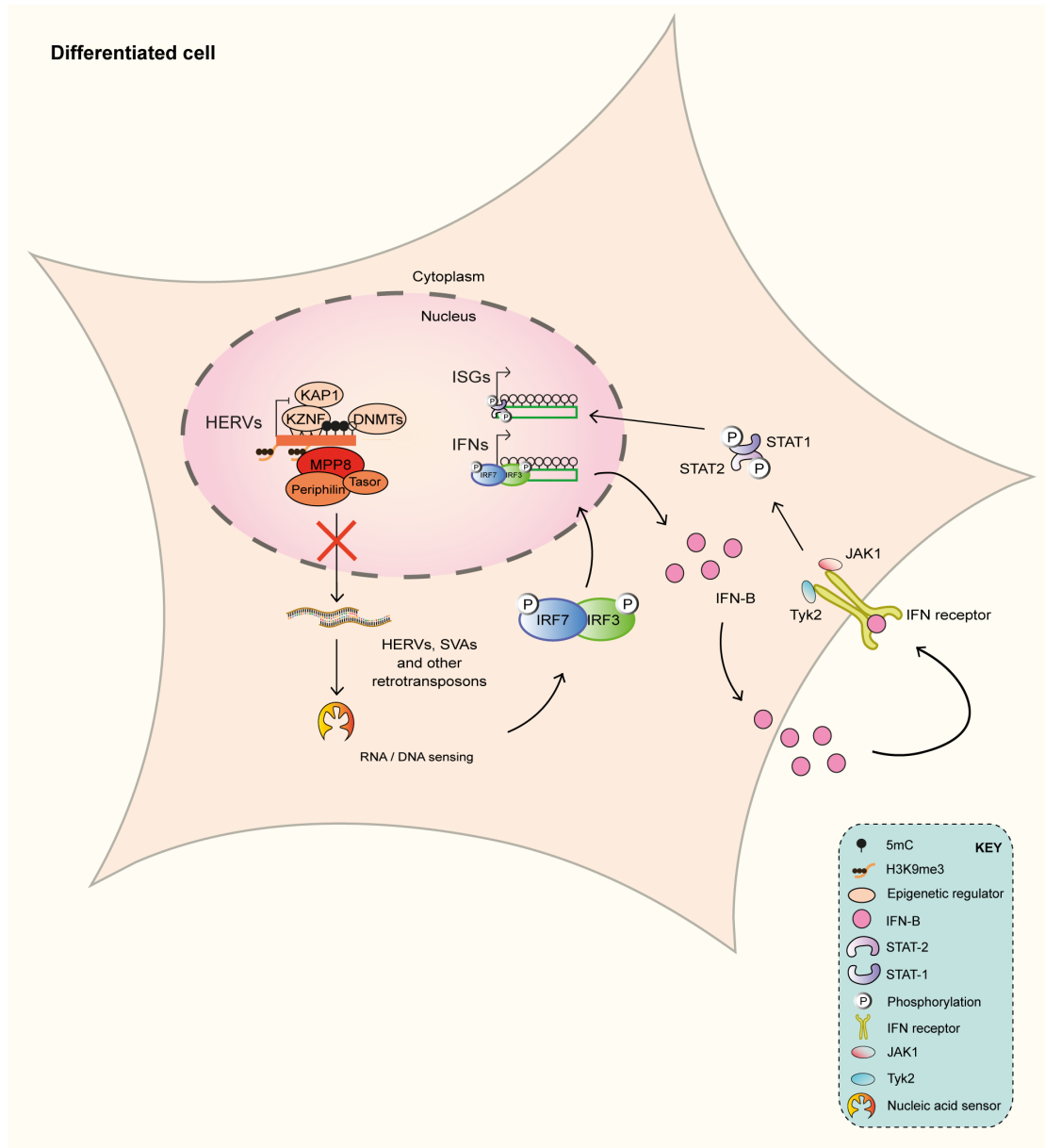


Figure 40 Epigenetic regulators protect cells from an unwanted innate immune response.

DNA methylation and potentially also H3K9me3 play a crucial role in protecting differentiated cells from innate immune activation elicited through viral mimicry. Here, we have identified MPP8 and to a lesser extent KAP1 as gatekeepers preventing immune reactivity. We also show that KAP1, SETDB1, the HUSH complex and DNA methylation act on some of the same retrotransposon sequences including on HERVK14C. ERVs may serve as PAMPs through their potential bidirectional transcription and production of double-stranded RNA and / or their cDNA synthesis through potential reverse transcription. Such PAMPs are well-known to trigger RNA and DNA sensors. Therefore, this nucleic acid sensing mimics a viral infection, which can trigger an innate and potentially also an adaptive immune response. We have detected the initial events of this response when we deplete MPP8, which involves classical type I IFN signalling.

both and whether MPP8 prevents viral mimicry through the maintenance of histone or DNA methylation or both? Finally, it is not clear if retrotransposon PAMPs play a role in the normal functioning of the immune system through beneficial innate immune activation in certain contexts.

Importantly, our data indicate that 5-Aza is not specific to cancer cells because it affects retrotransposons and ISGs in both cancer cell lines and primary PBMCs. Interestingly, it has previously been reported that the drug can also be used to induce regulatory T cells to inhibit autoimmune encephalomyelitis (Chan et al., 2014). Although we observed both SVA D and HERVK14C to be upregulated following 5-Aza treatment, it is unclear if either serve as PAMPs in the ISG response. Previous studies suggest that dsRNA from ERVs is central to ISG induction but the authors failed to directly show this relationship and it is possible that the upregulation of ERVs simply reflects the consequence of an inflammatory response in cancer cells (Chiappinelli et al., 2015; Roulois et al., 2015).

During viral infection, characteristic viral dsRNAs or 5'-triphosphate-ssRNAs (Jensen and Thomsen, 2012) act as PAMPs for RNA sensors. ERV transcripts, in contrast are transcribed and modified like cellular mRNAs. However, TEs can still give rise to dsRNA through different mechanisms. Firstly, some TEs have bidirectional promoters and can therefore produce dsRNAs (Sokol et al., 2015; Uesaka et al., 2014) as proposed (Chiappinelli et al., 2015; Roulois et al., 2015). Secondly, stem-loop structures within the RNA transcripts (potentially from SVAs) have been demonstrated to induce innate immune responses (Pichlmair et al., 2009; Sharma et al., 2011). The RNA editing protein, ADAR1, mediates the adenosine-to-inosine (A-to-I) editing of RNA transcripts and prevents the formation of stem loop structures

within endogenous transcripts in healthy cells (Aktas et al., 2017; Liddicoat et al., 2015). During DNA demethylation, an increase in ERV transcripts may potentially overwhelm the activity of ADAR1 leading to an accumulation of dsRNA and activation of RNA sensors.

The HDAC inhibitor, TSA, or KAP1 knockdown could not phenocopy the effects of 5-Aza and this may be because KAP1 knockout only significantly upregulates three families of ERVs, while 5-Aza treatment, affects more ERVs (Chiappinelli et al., 2015; Roulois et al., 2015), including SVAs. Similarly, TSA treatment might be insufficient in resurrecting most TEs due to the presence of repressive histone marks and DNA methylation as previously reported (Lynch et al., 2002; Svensson et al., 1998).

The lack of a synergistic or additive effect of KAP1 knockdown and drug treatment suggests that KAP1 exerts partial redundancy with other epigenetic pathways and that the level of redundancy at individual ERV integrants may relate to multiple parameters including their evolutionary age, local chromatin environment and nuclear position (Barklis et al., 1986; Jacobs et al., 2014; Reddy et al., 2008).

MPP8 is central to immune reactivity because its depletion does phenocopy 5-Aza treated cells and this is in line with the notion that many epigenetic regulators exert cross-talk with DNA methylation including SETDB1 and MPP8 (Chang et al., 2011; Karimi et al., 2011; Lee and Bieniasz, 2007). Overall in this chapter, we have shown that:

- i) Inhibiting DNMTs activates HERVK14C and SVA D elements while inhibiting HDACs activates HERVK14C and LIPA4

- ii) KAP1 knockdown can induce several chemokines but 5-Aza induces global ISGs in cancer cells and primary cells, potentially through SVAs
- iii) MPP8 knockdown phenocopies 5-Aza in ISG induction that proceeds through a classical IRF3 and JAK/STAT-dependent pathway

Chapter 6: Epilogue and future directions

My thesis has covered research into the role of KAP1 in regulating ERVs in differentiated cells, challenging the notion that the KZNF/KAP1 pathway is redundant in regulating ERVs beyond development. I will discuss my data in light of recent publications and speculate on the potential future developments of the project and field.

6.1 Significance of this thesis

My first aim (Chapters 3 and 4) was to investigate the ability of KAP1 to initiate and maintain ERV repression in differentiated cells. Most existing work in this field has been done in ES cells (Jacobs et al., 2014; Rowe et al., 2010; Turelli et al., 2014), whereas our data and parallel advancements in the field have all demonstrated a continuous role of KZNFs and KAP1 in the regulation of ERVs in diverse cell types (Brattas et al., 2017; Collins et al., 2015; Ecco et al., 2016). In support of a broad role for KAP1 in ERV regulation in adult tissues, we found KZNFs to be widely expressed, albeit at low levels. We have demonstrated here that these low levels are sufficient to induce *de novo* ERV repression both in cell lines and primary cells. KZNFs may be only lowly expressed due to lower activity of their target ERVs in somatic tissues. Alternatively, this observation might tie together with a recent report showing that ZFP809 (and potentially other KZNFs) are subject to protein degradation in differentiated cells (Wang and Goff, 2017). I hypothesise that KZNFs may function as restriction factors inhibiting endogenous as well as exogenous retroviruses and their expression might be induced in adult tissues, for example upon viral infection.

I anticipate more studies revealing the connection between ERVs and transcriptional regulation, particularly in the context of cancer, will be revealed in the coming years. With improvements in techniques used to study chromatin structure, more information will be uncovered on the role of ERVs in nuclear organisation too. Considering that ERV protein products can be expressed in adult tissues, I anticipate that another new area of research will involve mapping ERV interactomes.

In Chapter 5, I explored the link between ERVs and immune sensing. I showed that epigenetic factors, particularly MPP8, can protect the host from aberrant immune activation. Consistent with recent publications showing the ability of ERV transcripts to stimulate innate immune responses (Chiappinelli et al., 2015; Roulois et al., 2015; Zeng et al., 2014), my data point to a possible role of KAP1 and MPP8 epigenetic pathways in protecting immune integrity. These findings may lay the groundwork for further investigations on the interplay between epigenetics and innate immunity. ERVs likely play a broad role in the efficacy of cancer drugs through innate signalling (Chiappinelli et al., 2015; Goel et al., 2017; Roulois et al., 2015) and a better understanding of the mechanisms at play will be pivotal to the discovery of new drug targets for cancer treatments.

6.2 Future directions

Following my work, the next steps will be to explore the epigenetic changes at KAP1-regulated ERVs following KAP1-depletion in differentiated cells. Do these loci lose repressive histone marks or DNA methylation or both and gain active marks prior to their reactivation? I think it would also be interesting to test the hypothesis

that KZNFs function as restriction factors and dissect their potential role during viral infections. In addition, I would further investigate the mechanism responsible for the ISG response observed in Chapter 5. The primary questions that remain here are:

- i) What is the PAMP in 5-Aza treated, KAP1-depleted and MPP8-depleted cells? One potential candidate is SVA elements because of the secondary structure of their RNA.
- ii) What is the type of PAMP e.g. dsRNA, or cytoplasmic DNA or both and which PRRs are involved?
- iii) Does MPP8 depletion induce an antiviral state that is sufficient to restrict infections from exogenous viruses?

Understanding the link between the immune system and ERVs is particularly relevant to cancer therapy where patients are under 5-Aza-based treatments and can add an extra piece to the puzzle of ERV expression in patients with autoimmune diseases.

References

1. Ablasser, A., Hemmerling, I., Schmid-Burgk, J.L., Behrendt, R., Roers, A., and Hornung, V. (2014). TREX1 deficiency triggers cell-autonomous immunity in a cGAS-dependent manner. *J Immunol* 192, 5993-5997.
2. Aktas, T., Avsar Ilik, I., Maticzka, D., Bhardwaj, V., Pessoa Rodrigues, C., Mittler, G., Manke, T., Backofen, R., and Akhtar, A. (2017). DHX9 suppresses RNA processing defects originating from the Alu invasion of the human genome. *Nature* 544, 115-119.
3. Almine, J.F., O'Hare, C.A., Dunphy, G., Haga, I.R., Naik, R.J., Atrih, A., Connolly, D.J., Taylor, J., Kelsall, I.R., Bowie, A.G., *et al.* (2017). IFI16 and cGAS cooperate in the activation of STING during DNA sensing in human keratinocytes. *Nat Commun* 8, 14392.
4. Ambrosini, G., Dreos, R., Kumar, S., and Bucher, P. (2016). The ChIP-Seq tools and web server: a resource for analyzing ChIP-seq and other types of genomic data. *BMC Genomics* 17, 938.
5. Aravin, A.A., Hannon, G.J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318, 761-764.
6. Ariza, M.E., and Williams, M.V. (2011). A human endogenous retrovirus K dUTPase triggers a TH1, TH17 cytokine response: does it have a role in psoriasis? *J Invest Dermatol* 131, 2419-2427.
7. Ayyanathan, K., Lechner, M.S., Bell, P., Maul, G.G., Schultz, D.C., Yamada, Y., Tanaka, K., Torigoe, K., and Rauscher, F.J., 3rd (2003). Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation. *Genes Dev* 17, 1855-1869.
8. Babaian, A., and Mager, D.L. (2016). Endogenous retroviral promoter exaptation in human cancer. *Mob DNA* 7.
9. Babushok, D.V., and Kazazian, H.H., Jr. (2007). Progress in understanding the biology of the human mutagen LINE-1. *Hum Mutat* 28, 527-539.
10. Balada, E., Ordi-Ros, J., and Vilardell-Tarres, M. (2009). Molecular mechanisms mediated by human endogenous retroviruses (HERVs) in autoimmunity. *Rev Med Virol* 19, 273-286.
11. Banki, K., Maceda, J., Hurley, E., Ablonczy, E., Mattson, D.H., Szegedy, L., Hung, C., and Perl, A. (1992). Human T-cell lymphotropic virus (HTLV)-related endogenous sequence, HRES-1, encodes a 28-kDa protein: a possible autoantigen for HTLV-I gag-reactive autoantibodies. *Proc Natl Acad Sci U S A* 89, 1939-1943.
12. Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410, 120-124.
13. Barklis, E., Mulligan, R.C., and Jaenisch, R. (1986). Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. *Cell* 47, 391-399.
14. Barrat, F.J., Meeker, T., Gregorio, J., Chan, J.H., Uematsu, S., Akira, S., Chang, B., Duramad, O., and Coffman, R.L. (2005). Nucleic acids of

mammalian origin can act as endogenous ligands for Toll-like receptors and may promote systemic lupus erythematosus. *J Exp Med* 202, 1131-1139.

15. Batzer, M.A., and Deininger, P.L. (2002). Alu repeats and human genomic diversity. *Nat Rev Genet* 3, 370-379.
16. Baudino, L., Yoshinobu, K., Dunand-Sauthier, I., Evans, L.H., and Izui, S. (2010). TLR-mediated up-regulation of serum retroviral gp70 is controlled by the Sgp loci of lupus-prone mice. *J Autoimmun* 35, 153-159.
17. Beck, C.R., Garcia-Perez, J.L., Badge, R.M., and Moran, J.V. (2011). LINE-1 elements in structural variation and disease. *Annu Rev Genomics Hum Genet* 12, 187-215.
18. Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G., Pačes, J., Burt, A., and Tristem, M. (2004). Long-term reinfection of the human genome by endogenous retroviruses. In *Proc Natl Acad Sci U S A*, pp. 4894-4899.
19. Belshaw, R., Watson, J., Katzourakis, A., Howe, A., Woolven-Allen, J., Burt, A., and Tristem, M. (2007). Rate of Recombinational Deletion among Human Endogenous Retroviruses. In *J Virol*, pp. 9437-9442.
20. Best, S., Le Tissier, P., Towers, G., and Stoye, J.P. (1996). Positional cloning of the mouse retrovirus restriction gene Fv1. *Nature* 382, 826-829.
21. Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988). Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol* 203, 971-983.
22. Bestor, T.H. (1988). Cloning of a mammalian DNA methyltransferase. *Gene* 74, 9-12.
23. Beyer, U., Moll-Rocek, J., Moll, U.M., and Dobbelstein, M. (2011). Endogenous retrovirus drives hitherto unknown proapoptotic p63 isoforms in the male germ line of humans and great apes. In *Proc Natl Acad Sci U S A*, pp. 3624-3629.
24. Bieda, K., Hoffmann, A., and Boller, K. (2001). Phenotypic heterogeneity of human endogenous retrovirus particles produced by teratocarcinoma cell lines. *J Gen Virol* 82, 591-596.
25. Blaise, S., de Parseval, N., Benit, L., and Heidmann, T. (2003). Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc Natl Acad Sci U S A* 100, 13013-13018.
26. Blanco-Melo, D., Gifford, R.J., and Bieniasz, P.D. (2017). Co-option of an endogenous retrovirus envelope for host defense in hominid ancestors. *Elife* 6.
27. Blomberg, J., Benachenhou, F., Blikstad, V., Sperber, G., and Mayer, J. (2009). Classification and nomenclature of endogenous retroviral sequences (ERVs): problems and recommendations. *Gene* 448, 115-123.
28. Bock, M., and Stoye, J.P. (2000). Endogenous retroviruses and the human germline. *Curr Opin Genet Dev* 10, 651-655.
29. Boeke, J., and Stoye, J. (1997). Retrotransposons, Endogenous Retroviruses, and the Evolution of Retroelements. 343-436.

30. Boissinot, S., Chevret, P., and Furano, A.V. (2000). L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol Biol Evol* 17, 915-928.
31. Boissinot, S., and Furano, A.V. (2001). Adaptive evolution in LINE-1 retrotransposons. *Mol Biol Evol* 18, 2186-2194.
32. Bojkowska, K., Aloisio, F., Cassano, M., Kapopoulou, A., Santoni de Sio, F., Zangger, N., Offner, S., Cartoni, C., Thomas, C., Quenneville, S., *et al.* (2012). Liver-specific ablation of Kruppel-associated box-associated protein 1 in mice leads to male-predominant hepatosteatosis and development of liver adenoma. *Hepatology* 56, 1279-1290.
33. Brattas, P.L., Jonsson, M.E., Fasching, L., Nelander Wahlestedt, J., Shahsavani, M., Falk, R., Falk, A., Jern, P., Parmar, M., and Jakobsson, J. (2017). TRIM28 Controls a Gene Regulatory Network Based on Endogenous Retroviruses in Human Neural Progenitor Cells. *Cell Rep* 18, 1-11.
34. Britten, R.J., and Davidson, E.H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev Biol* 46, 111-138.
35. Bronkhorst, A.W., and van Rij, R.P. (2014). The long and short of antiviral defense: small RNA-based immunity in insects. *Curr Opin Virol* 7, 19-28.
36. Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., and Kazazian, H.H., Jr. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* 100, 5280-5285.
37. Browne, E.P. (2011). Toll-like receptor 7 controls the anti-retroviral germinal center response. *PLoS Pathog* 7, e1002293.
38. Bulut-Karslioglu, A., De La Rosa-Velazquez, I.A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J., Galan, C., Winter, G.E., Engist, B., Gerle, B., *et al.* (2014). Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol Cell* 55, 277-290.
39. Buscher, K., Trefzer, U., Hofmann, M., Sterry, W., Kurth, R., and Denner, J. (2005). Expression of human endogenous retrovirus K in melanomas and melanoma cell lines. *Cancer Res* 65, 4172-4180.
40. Cammas, F., Herzog, M., Lerouge, T., Chambon, P., and Losson, R. (2004). Association of the transcriptional corepressor TIF1beta with heterochromatin protein 1 (HP1): an essential role for progression through differentiation. *Genes Dev* 18, 2147-2160.
41. Cammas, F., Mark, M., Dolle, P., Dierich, A., Chambon, P., and Losson, R. (2000). Mice lacking the transcriptional corepressor TIF1beta are defective in early postimplantation development. *Development* 127, 2955-2963.
42. Canzio, D., Liao, M., Naber, N., Pate, E., Larson, A., Wu, S., Marina, D.B., Garcia, J.F., Madhani, H.D., Cooke, R., *et al.* (2013). A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. *Nature* 496, 377-381.
43. Chan, M.W.Y., Chang, C.B., Tung, C.H., Sun, J., Suen, J.L., and Wu, S.F. (2014). Low-Dose 5-Aza-2'-deoxycytidine Pretreatment Inhibits

Experimental Autoimmune Encephalomyelitis by Induction of Regulatory T Cells. In *Mol Med*, pp. 248-256.

44. Chang, C.J., Chen, Y.L., and Lee, S.C. (1998). Coactivator TIF1beta interacts with transcription factor C/EBPbeta and glucocorticoid receptor to induce alpha1-acid glycoprotein gene expression. *Mol Cell Biol* 18, 5880-5887.
45. Chang, Y., Sun, L., Kokura, K., Horton, J.R., Fukuda, M., Espejo, A., Izumi, V., Koomen, J.M., Bedford, M.T., Zhang, X., *et al.* (2011). MPP8 mediates the interactions between DNA methyltransferase Dnmt3a and H3K9 methyltransferase GLP/G9a. *Nat Commun* 2, 533.
46. Chedin, F., Lieber, M.R., and Hsieh, C.L. (2002). The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a. *Proc Natl Acad Sci U S A* 99, 16916-16921.
47. Chen, L., Dahlstrom, J.E., Lee, S.H., and Rangasamy, D. (2012). Naturally occurring endo-siRNA silences LINE-1 retrotransposons in human cells through DNA methylation. *Epigenetics* 7, 758-771.
48. Chen, T., Meng, Z., Gan, Y., Wang, X., Xu, F., Gu, Y., Xu, X., Tang, J., Zhou, H., Zhang, X., *et al.* (2013). The viral oncogene Np9 acts as a critical molecular switch for co-activating beta-catenin, ERK, Akt and Notch1 and promoting the growth of human leukemia stem/progenitor cells. *Leukemia* 27, 1469-1478.
49. Chen, W.Y., and Townes, T.M. (2000). Molecular mechanism for silencing virally transduced genes involves histone deacetylation and chromatin condensation. In *Proc Natl Acad Sci U S A*, pp. 377-382.
50. Chiappinelli, K.B., Strissel, P.L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N.S., Cope, L.M., Snyder, A., *et al.* (2015). Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* 162, 974-986.
51. Chow, J.C., Ciaudo, C., Fazzari, M.J., Mise, N., Servant, N., Glass, J.L., Attreed, M., Avner, P., Wutz, A., Barillot, E., *et al.* (2016). LINE-1 Activity in Facultative Heterochromatin Formation during X Chromosome Inactivation. *Cell* 166, 782.
52. Christensen, T. (2016). Human endogenous retroviruses in neurologic disease. *Apmis* 124, 116-126.
53. Chuong, E.B., Elde, N.C., and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351, 1083-1087.
54. Chuong, E.B., Rumi, M.A., Soares, M.J., and Baker, J.C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* 45, 325-329.
55. Collins, P.L., Kyle, K.E., Egawa, T., Shinkai, Y., and Oltz, E.M. (2015). The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes. *Proc Natl Acad Sci U S A* 112, 8367-8372.
56. Comeaux, M.S., Roy-Engel, A.M., Hedges, D.J., and Deininger, P.L. (2009). Diverse cis factors controlling Alu retrotransposition: what causes Alu elements to die? *Genome Res* 19, 545-555.
57. Conrad, B., Weidmann, E., Trucco, G., Rudert, W.A., Behboo, R., Ricordi, C., Rodriquez-Rilo, H., Finegold, D., and Trucco, M. (1994). Evidence for

superantigen involvement in insulin-dependent diabetes mellitus aetiology. *Nature* 371, 351-355.

58. Conrad, B., Weissmahr, R.N., Boni, J., Arcari, R., Schupbach, J., and Mach, B. (1997). A human endogenous retroviral superantigen as candidate autoimmune gene in type I diabetes. *Cell* 90, 303-313.
59. Consortium, C.S.a.A. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437, 69-87.
60. Cordaux, R., and Batzer, M.A. (2009). The impact of retrotransposons on human genome evolution. *Nature reviews Genetics* 10, 691-703.
61. Cornelis, G., Vernochet, C., Carradec, Q., Souquere, S., Mulot, B., Catzeflis, F., Nilsson, M.A., Menzies, B.R., Renfree, M.B., Pierron, G., *et al.* (2015). Retroviral envelope gene captures and syncytin exaptation for placentation in marsupials. *Proc Natl Acad Sci U S A* 112, E487-496.
62. Corsinotti, A., Kapopoulou, A., Gubelmann, C., Imbeault, M., Santoni de Sio, F.R., Rowe, H.M., Mouscaz, Y., Deplancke, B., and Trono, D. (2013). Global and stage specific patterns of Kruppel-associated-box zinc finger protein gene expression in murine early embryonic cells. *PLoS One* 8, e56721.
63. Cost, G.J., Feng, Q., Jacquier, A., and Boeke, J.D. (2002). Human L1 element target-primed reverse transcription in vitro. In *EMBO J*, pp. 5899-5910.
64. Craigie, R., and Bushman, F.D. (2012). HIV DNA integration. *Cold Spring Harb Perspect Med* 2, a006890.
65. Daniels, G.R., and Deininger, P.L. (1985). Repeat sequence families derived from mammalian tRNA genes. *Nature* 317, 819-822.
66. de Koning, A.P., Gu, W., Castoe, T.A., Batzer, M.A., and Pollock, D.D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 7, e1002384.
67. Dennis, K., Fan, T., Geiman, T., Yan, Q., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev* 15, 2940-2944.
68. Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 35, 41-48.
69. Dong, K.B., Maksakova, I.A., Mohn, F., Leung, D., Appanah, R., Lee, S., Yang, H.W., Lam, L.L., Mager, D.L., Schubeler, D., *et al.* (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *Embo j* 27, 2691-2701.
70. Douville, R., Liu, J., Rothstein, J., and Nath, A. (2011). Identification of active loci of a human endogenous retrovirus in neurons of patients with amyotrophic lateral sclerosis. *Ann Neurol* 69, 141-151.
71. Douville, R.N., and Nath, A. (2014). Human endogenous retroviruses and the nervous system. *Handb Clin Neurol* 123, 465-485.
72. Dunn, C.A., Romanish, M.T., Gutierrez, L.E., van de Lagemaat, L.N., and Mager, D.L. (2006). Transcription of two human genes from a bidirectional endogenous retrovirus promoter. *Gene* 366, 335-342.
73. Dupressoir, A., Marceau, G., Vernochet, C., Bénit, L., Kanellopoulos, C., Sapin, V., and Heidmann, T. (2005). Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. In *Proc Natl Acad Sci U S A*, pp. 725-730.

74. Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P., and Heidmann, T. (2009). Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc Natl Acad Sci U S A* 106, 12127-12132.
75. Durruthy-Durruthy, J., Sebastiano, V., Wossidlo, M., Cepeda, D., Cui, J., Grow, E.J., Davila, J., Mall, M., Wong, W.H., Wysocka, J., *et al.* (2016). The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming. *Nat Genet* 48, 44-52.
76. Ebert, P.J., Jiang, S., Xie, J., Li, Q.J., and Davis, M.M. (2009). An endogenous positively selecting peptide enhances mature T cell responses and becomes an autoantigen in the absence of microRNA miR-181a. *Nat Immunol* 10, 1162-1169.
77. Ecco, G., Cassano, M., Kauzlaric, A., Duc, J., Coluccio, A., Offner, S., Imbeault, M., Rowe, H.M., Turelli, P., and Trono, D. (2016). Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues. *Dev Cell* 36, 611-623.
78. Eckard, S.C., Rice, G.I., Fabre, A., Badens, C., Gray, E.E., Hartley, J.L., Crow, Y.J., and Stetson, D.B. (2014). The SKIV2L RNA exosome limits activation of the RIG-I-like receptors. *Nat Immunol* 15, 839-845.
79. Ehlhardt, S., Seifert, M., Schneider, J., Ojak, A., Zang, K.D., and Mehraein, Y. (2006). Human endogenous retrovirus HERV-K(HML-2) Rec expression and transcriptional activities in normal and rheumatoid arthritis synovia. *J Rheumatol* 33, 16-23.
80. Ejtehadi, H.D., Freimanis, G.L., Ali, H.A., Bowman, S., Alavi, A., Axford, J., Callaghan, R., and Nelson, P.N. (2006). The potential role of human endogenous retrovirus K10 in the pathogenesis of rheumatoid arthritis: a preliminary study. *Ann Rheum Dis* 65, 612-616.
81. Elsässer, S.J., Noh, K.-M., Diaz, N., Allis, C.D., and Banaszynski, L.A. (2015). Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature* 522, 240-244.
82. Emera, D., Casola, C., Lynch, V.J., Wildman, D.E., Agnew, D., and Wagner, G.P. (2012). Convergent evolution of endometrial prolactin expression in primates, mice, and elephants through the independent recruitment of transposable elements. *Mol Biol Evol* 29, 239-247.
83. Emerson, R.O., and Thomas, J.H. (2009). Adaptive evolution in zinc finger transcription factors. *PLoS Genet* 5, e1000325.
84. Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A.J., Heidmann, T., and Schwartz, O. (2005). APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* 433, 430-433.
85. Fasching, L., Kapopoulou, A., Sachdeva, R., Petri, R., Jonsson, M.E., Manne, C., Turelli, P., Jern, P., Cammas, F., Trono, D., *et al.* (2015). TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. *Cell Rep* 10, 20-28.
86. Feng, Q., Moran, J.V., Kazazian, H.H., Jr., and Boeke, J.D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87, 905-916.

87. Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. *Science* 330, 622-627.
88. Fernandez, D.R., Telarico, T., Bonilla, E., Li, Q., Banerjee, S., Middleton, F.A., Phillips, P.E., Crow, M.K., Oess, S., Muller-Esterl, W., *et al.* (2009). Activation of mammalian target of rapamycin controls the loss of TCRzeta in lupus T cells through HRES-1/Rab4-regulated lysosomal degradation. *J Immunol* 182, 2063-2073.
89. Feschotte, C., and Gilbert, C. (2012). Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet* 13, 283-296.
90. Feuchter, A., and Mager, D. (1990). Functional heterogeneity of a large family of human LTR-like promoters and enhancers. *Nucleic Acids Res* 18, 1261-1270.
91. Fitch, W.M. (1987). Of urfs and orfs: A primer on how to analyze derived amino acid sequences. *Cell* 50, 329.
92. Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C.A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A., *et al.* (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat Genet* 46, 558-566.
93. Friedli, M., and Trono, D. (2015). The developmental control of transposable elements and the evolution of higher species. *Annu Rev Cell Dev Biol* 31, 429-451.
94. Friedli, M., Turelli, P., Kapopoulou, A., Rauwel, B., Castro-Diaz, N., Rowe, H.M., Ecco, G., Unzu, C., Planet, E., Lombardo, A., *et al.* (2014). Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res* 24, 1251-1259.
95. Frietze, S., O'Geen, H., Blahnik, K.R., Jin, V.X., and Farnham, P.J. (2010). ZNF274 Recruits the Histone Methyltransferase SETDB1 to the 3' Ends of ZNF Genes. *PLoS One* 5, e15082.
96. Furano, A.V., USA, N.I.o.H.B.M., Boissinot, S., USA, N.I.o.H.B.M., Furano, A.V., and Boissinot, S. (2006). Long Interspersed Nuclear Elements (LINEs): Evolution (John Wiley & Sons, Ltd).
97. Gao, D., Wu, J., Wu, Y.-T., Du, F., Aroh, C., Yan, N., Sun, L., and Chen, Z.J. (2013). Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. *Science (New York, NY)* 341, 903-906.
98. Gao, Y., Meng, H., Liu, S., Hu, J., Zhang, Y., Jiao, T., Liu, Y., Ou, J., Wang, D., Yao, L., *et al.* (2015). LncRNA-HOST2 regulates cell biological behaviors in epithelial ovarian cancer through a mechanism involving microRNA let-7b. *Hum Mol Genet* 24, 841-852.
99. Garcia-Perez, J.L., Morell, M., Scheys, J.O., Kulpa, D.A., Morell, S., Carter, C.C., Hammer, G.D., Collins, K.L., O'Shea, K.S., Menendez, P., *et al.* (2010). Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* 466, 769-773.
100. Gardner, M.B., Kozak, C.A., and O'Brien, S.J. (1991). The Lake Casitas wild mouse: evolving genetic resistance to retroviral disease. *Trends Genet* 7, 22-27.
101. Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. *Nat Rev Genet* 10, 94-108.

102. Gifford, R., and Tristem, M. (2003). The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* 26, 291-315.
103. Gifford, R.J., Katzourakis, A., Tristem, M., Pybus, O.G., Winters, M., and Shafer, R.W. (2008). A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proc Natl Acad Sci U S A* 105, 20362-20367.
104. Gifford, W.D., Pfaff, S.L., and Macfarlan, T.S. (2013). Transposable elements as genetic regulatory substrates in early development. *Trends Cell Biol* 23, 218-226.
105. Girard, A., Sachidanandam, R., Hannon, G.J., and Carmell, M.A. (2006). A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature* 442, 199-202.
106. Goel, S., DeCristo, M.J., Watt, A.C., BrinJones, H., Sceneay, J., Li, B.B., Khan, N., Ubellacker, J.M., Xie, S., Metzger-Filho, O., *et al.* (2017). CDK4/6 inhibition triggers anti-tumour immunity. *Nature* 548, 471-475.
107. Goering, W., Schmitt, K., Dostert, M., Schaal, H., Deenen, R., Mayer, J., and Schulz, W.A. (2015). Human endogenous retrovirus HERV-K(HML-2) activity in prostate cancer is dominated by a few loci. *Prostate* 75, 1958-1971.
108. Goke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F., and Szczerbinska, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* 16, 135-141.
109. Golan, M., Hizi, A., Resau, J.H., Yaal-Hahoshen, N., Reichman, H., Keydar, I., and Tsarfaty, I. (2008). Human endogenous retrovirus (HERV-K) reverse transcriptase as a breast cancer prognostic marker. *Neoplasia* 10, 521-533.
110. Golding, M.C., Zhang, L., and Mann, M.R. (2010). Multiple epigenetic modifiers induce aggressive viral extinction in extraembryonic endoderm stem cells. *Cell Stem Cell* 6, 457-467.
111. Goodier, J.L., and Kazazian, H.H., Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* 135, 23-35.
112. Graham, F.L., Smiley, J., Russell, W.C., and Nairn, R. (1977). Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J Gen Virol* 36, 59-74.
113. Griffiths, D.J. (2001). Endogenous retroviruses in the human genome sequence. *Genome Biol* 2, 1017.
114. Grimaldi, G., Skowronski, J., and Singer, M.F. (1984). Defining the beginning and end of KpnI family segments. *EMBO J* 3, 1753-1759.
115. Groner, A.C., Meylan, S., Ciuffi, A., Zangger, N., Ambrosini, G., Denervaud, N., Bucher, P., and Trono, D. (2010). KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLoS Genet* 6, e1000869.
116. Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J., Martin, L., Ware, C.B., Blish, C.A., Chang, H.Y., *et al.* (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 522, 221-225.
117. Gürtler, C., and Bowie, A.G. (2013). Innate immune detection of microbial nucleic acids. *Trends Microbiol* 21, 413-420.

118. Hardies, S.C., Martin, S.L., Voliva, C.F., Hutchison, C.A., 3rd, and Edgell, M.H. (1986). An analysis of replacement and synonymous changes in the rodent L1 repeat family. *Mol Biol Evol* 3, 109-125.
119. Hayashi, M., Maehara, K., Harada, A., Semba, Y., Kudo, K., Takahashi, H., Oki, S., Meno, C., Ichiyanagi, K., Akashi, K., *et al.* (2016). Chd5 Regulates MuERV-L/MERVL Expression in Mouse Embryonic Stem Cells Via H3K27me3 Modification and Histone H3.1/H3.2. *J Cell Biochem* 117, 780-792.
120. Hayward, A. (2017). Origin of the retroviruses: when, where, and how? *Curr Opin Virol* 25, 23-27.
121. Hegyi, H. (2013). GABBR1 has a HERV-W LTR in its regulatory region – a possible implication for schizophrenia. In *Biol Direct*, p. 5.
122. Herrmann, M., Neidhart, M., Gay, S., Hagenhofer, M., and Kalden, J.R. (1998). Retrovirus-associated rheumatic syndromes. *Curr Opin Rheumatol* 10, 347-354.
123. Herve, C.A., Lugli, E.B., Brand, A., Griffiths, D.J., and Venables, P.J. (2002). Autoantibodies to human endogenous retrovirus-K are frequently detected in health and disease and react with multiple epitopes. *Clin Exp Immunol* 128, 75-82.
124. Hiragami-Hamada, K., Soeroes, S., Nikolov, M., Wilkins, B., Kreuz, S., Chen, C., De La Rosa-Velazquez, I.A., Zenn, H.M., Kost, N., Pohl, W., *et al.* (2016). Dynamic and flexible H3K9me3 bridging via HP1beta dimerization establishes a plastic state of condensed chromatin. *Nat Commun* 7, 11310.
125. Hisada, K., Sanchez, C., Endo, T.A., Endoh, M., Roman-Trufero, M., Sharif, J., Koseki, H., and Vidal, M. (2012). RYBP represses endogenous retroviruses and preimplantation- and germ line-specific genes in mouse embryonic stem cells. *Mol Cell Biol* 32, 1139-1149.
126. Hiscott, J. (2007). Triggering the innate antiviral response through IRF-3 activation. *J Biol Chem* 282, 15325-15329.
127. Hohjoh, H., and Singer, M.F. (1996). Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *Embo j* 15, 630-639.
128. Hohn, O., Hanke, K., and Bannert, N. (2013). HERV-K(HML-2), the Best Preserved Family of HERVs: Endogenization, Expression, and Implications in Health and Disease. *Front Oncol* 3, 246.
129. Holmes, S.E., Dombroski, B.A., Krebs, C.M., Boehm, C.D., and Kazazian, H.H., Jr. (1994). A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nat Genet* 7, 143-148.
130. Hornung, V., Ablasser, A., Charrel-Dennis, M., Bauernfeind, F., Horvath, G., Caffrey, D.R., Latz, E., and Fitzgerald, K.A. (2009). AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature* 458, 514-518.
131. Hu, W.S., and Hughes, S.H. (2012). HIV-1 reverse transcription. *Cold Spring Harb Perspect Med* 2, a006882.
132. Huang, J., Fan, T., Yan, Q., Zhu, H., Fox, S., Issaq, H.J., Best, L., Gangi, L., Munroe, D., and Muegge, K. (2004). Lsh, an epigenetic guardian of repetitive elements. *Nucleic Acids Res* 32, 5019-5028.

133. Hummel, B., Hansen, E.C., Yoveva, A., Aprile-Garcia, F., Hussong, R., and Sawarkar, R. (2017). The evolutionary capacitor HSP90 buffers the regulatory effects of mammalian endogenous retroviruses. *Nat Struct Mol Biol* 24, 234-242.
134. Huntley, S., Baggott, D.M., Hamilton, A.T., Tran-Gyamfi, M., Yang, S., Kim, J., Gordon, L., Branscomb, E., and Stubbs, L. (2006). A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. *Genome Res* 16, 669-677.
135. Ichiyanagi, K. (2013). Epigenetic regulation of transcription and possible functions of mammalian short interspersed elements, SINEs. *Genes Genet Syst* 88, 19-29.
136. Imbeault, M., Helleboid, P.Y., and Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* 543, 550-554.
137. Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Boskovic, A., Ziegler-Birling, C., Rodriguez-Terrones, D., Wakayama, T., Vaquerizas, J.M., and Torres-Padilla, M.E. (2015). Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat Struct Mol Biol* 22, 662-671.
138. Ivanov, A.V., Peng, H., Yurchenko, V., Yap, K.L., Negorev, D.G., Schultz, D.C., Psulkowski, E., Fredericks, W.J., White, D.E., Maul, G.G., *et al.* (2007). PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol Cell* 28, 823-837.
139. Iyengar, S., and Farnham, P.J. (2011). KAP1 protein: an enigmatic master regulator of the genome. *J Biol Chem* 286, 26267-26276.
140. Iyengar, S., Ivanov, A.V., Jin, V.X., Rauscher, F.J., 3rd, and Farnham, P.J. (2011). Functional analysis of KAP1 genomic recruitment. *Mol Cell Biol* 31, 1833-1847.
141. Jacobs, F.M., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A.D., Katzman, S., Paten, B., Salama, S.R., and Haussler, D. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* 516, 242-245.
142. Jacques, P.E., Jeyakani, J., and Bourque, G. (2013). The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet* 9, e1003504.
143. Jahner, D., Stuhlmann, H., Stewart, C.L., Harbers, K., Lohler, J., Simon, I., and Jaenisch, R. (1982). De novo methylation and expression of retroviral genomes during mouse embryogenesis. *Nature* 298, 623-628.
144. Jakobsen, M.R., Bak, R.O., Andersen, A., Berg, R.K., Jensen, S.B., Tengchuan, J., Laustsen, A., Hansen, K., Ostergaard, L., Fitzgerald, K.A., *et al.* (2013). IFI16 senses DNA forms of the lentiviral replication cycle and controls HIV-1 replication. *Proc Natl Acad Sci U S A* 110, E4571-4580.
145. Jensen, S., and Thomsen, A.R. (2012). Sensing of RNA viruses: a review of innate immune receptors involved in recognizing RNA virus invasion. *J Virol* 86, 2900-2910.
146. Jern, P., and Coffin, J.M. (2008). Effects of retroviruses on host genome function. *Annu Rev Genet* 42, 709-732.

147. Jern, P., Sperber, G.O., and Blomberg, J. (2005). Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. *Retrovirology* 2, 50.
148. Johnston, J.B., Silva, C., Holden, J., Warren, K.G., Clark, A.W., and Power, C. (2001). Monocyte activation and differentiation augment human endogenous retrovirus expression: implications for inflammatory brain diseases. *Ann Neurol* 50, 434-442.
149. Jones, P.L., Veenstra, G.J., Wade, P.A., Vermaak, D., Kass, S.U., Landsberger, N., Strouboulis, J., and Wolffe, A.P. (1998). Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 19, 187-191.
150. Kaiser, S.M., Malik, H.S., and Emerman, M. (2007). Restriction of an extinct retrovirus by the human TRIM5 α antiviral protein. *Science* 316, 1756-1758.
151. Kannan, S., Chernikova, D., Rogozin, I.B., Poliakov, E., Managadze, D., Koonin, E.V., and Milanese, L. (2015). Transposable Element Insertions in Long Intergenic Non-Coding RNA Genes. *Front Bioeng Biotechnol* 3.
152. Kapitonov, V.V., and Jurka, J. (2003). A novel class of SINE elements derived from 5S rRNA. *Mol Biol Evol* 20, 694-702.
153. Kapusta, A., Kronenberg, Z., Lynch, V.J., Zhuo, X., Ramsay, L., Bourque, G., Yandell, M., and Feschotte, C. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet* 9, e1003470.
154. Karimi, M.M., Goyal, P., Maksakova, I.A., Bilenky, M., Leung, D., Tang, J.X., Shinkai, Y., Mager, D.L., Jones, S., Hirst, M., *et al.* (2011). DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs. *Cell Stem Cell* 8, 676-687.
155. Kassiotis, G. (2014). Endogenous retroviruses and the development of cancer. *J Immunol* 192, 1343-1349.
156. Kato, H., Takeuchi, O., Mikamo-Satoh, E., Hirai, R., Kawai, T., Matsushita, K., Hiiragi, A., Dermody, T.S., Fujita, T., and Akira, S. (2008). Length-dependent recognition of double-stranded ribonucleic acids by retinoic acid-inducible gene-I and melanoma differentiation-associated gene 5. *J Exp Med* 205, 1601-1610.
157. Katoh, I., and Kurata, S. (2013). Association of Endogenous Retroviruses and Long Terminal Repeats with Human Disorders. *Front Oncol* 3.
158. Katzourakis, A., Pereira, V., and Tristem, M. (2007a). Effects of Recombination Rate on Human Endogenous Retrovirus Fixation and Persistence. In *J Virol*, pp. 10712-10717.
159. Katzourakis, A., Tristem, M., Pybus, O.G., and Gifford, R.J. (2007b). Discovery and analysis of the first endogenous lentivirus. *Proc Natl Acad Sci U S A* 104, 6261-6265.
160. Kaufmann, S., Sauter, M., Schmitt, M., Baumert, B., Best, B., Boese, A., Roemer, K., and Mueller-Lantzsch, N. (2010). Human endogenous retrovirus protein Rec interacts with the testicular zinc-finger protein and androgen receptor. *J Gen Virol* 91, 1494-1502.
161. Kauzlaric, A., Ecco, G., Cassano, M., Duc, J., Imbeault, M., and Trono, D. (2017). The mouse genome displays highly dynamic populations of

- KRAB-zinc finger protein genes and related genetic units. *PLoS One* 12, e0173746.
162. Kawai, T., and Akira, S. (2009). The roles of TLRs, RLRs and NLRs in pathogen recognition. *Int Immunol* 21, 317-337.
 163. Kazazian, H.H., Jr., and Moran, J.V. (1998). The impact of L1 retrotransposons on the human genome. *Nat Genet* 19, 19-24.
 164. Kelley, D., and Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol* 13, R107.
 165. Kershaw, M.H., Hsu, C., Mondesire, W., Parker, L.L., Wang, G., Overwijk, W.W., Lapointe, R., Yang, J.C., Wang, R.F., Restifo, N.P., *et al.* (2001). Immunization against endogenous retroviral tumor-associated antigens. *Cancer Res* 61, 7920-7924.
 166. Kigami, D., Minami, N., Takayama, H., and Imai, H. (2003). MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol Reprod* 68, 651-654.
 167. Kim, S., Gunesdogan, U., Zylicz, J.J., Hackett, J.A., Cougot, D., Bao, S., Lee, C., Dietmann, S., Allen, G.E., Sengupta, R., *et al.* (2014). PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. *Mol Cell* 56, 564-579.
 168. Kim, T.K., Hemberg, M., and Gray, J.M. (2015). Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. *Cold Spring Harb Perspect Biol* 7, a018622.
 169. Komashko, V.M., and Farnham, P.J. (2010). 5-azacytidine treatment reorganizes genomic histone modification patterns. *Epigenetics* 5, 229-240.
 170. Kourmouli, N., Jeppesen, P., Mahadevhaiah, S., Burgoyne, P., Wu, R., Gilbert, D.M., Bongiorno, S., Prantera, G., Fanti, L., Pimpinelli, S., *et al.* (2004). Heterochromatin and tri-methylated lysine 20 of histone H4 in animals. *J Cell Sci* 117, 2491-2501.
 171. Kozak, C.A. (2015). Origins of the Endogenous and Infectious Laboratory Mouse Gammaretroviruses. In *Viruses*, pp. 1-26.
 172. Kramerov, D.A., and Vassetzky, N.S. (2011). Origin and evolution of SINEs in eukaryotic genomes. *Heredity (Edinb)* 107, 487-495.
 173. Kreher, S., Bouhrel, M.A., Cauchy, P., Lamprecht, B., Li, S., Grau, M., Hummel, F., Kochert, K., Anagnostopoulos, I., Johrens, K., *et al.* (2014). Mapping of transcription factor motifs in active chromatin identifies IRF5 as key regulator in classical Hodgkin lymphoma. *Proc Natl Acad Sci U S A* 111, E4513-4522.
 174. Krieg, A.M., Gourley, M.F., Klinman, D.M., Perl, A., and Steinberg, A.D. (1992). Heterogeneous expression and coordinate regulation of endogenous retroviral sequences in human peripheral blood mononuclear cells. *AIDS Res Hum Retroviruses* 8, 1991-1998.
 175. Kubo, S., Seleme, M.C., Soifer, H.S., Perez, J.L., Moran, J.V., Kazazian, H.H., Jr., and Kasahara, N. (2006). L1 retrotransposition in nondividing and primary human somatic cells. *Proc Natl Acad Sci U S A* 103, 8036-8041.
 176. Kulis, M., and Esteller, M. (2010). DNA methylation and cancer. *Adv Genet* 70, 27-56.

177. Kulpa, D.A., and Moran, J.V. (2005). Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum Mol Genet* *14*, 3237-3248.
178. Kulpa, D.A., and Moran, J.V. (2006). Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat Struct Mol Biol* *13*, 655-660.
179. Kuznetsova, I.S., Ostromyshenskii, D.I., Komissarov, A.S., Prusov, A.N., Waisertreiger, I.S., Gorbunova, A.V., Trifonov, V.A., Ferguson-Smith, M.A., and Podgornaya, O.I. (2016). LINE-related component of mouse heterochromatin and complex chromocenters' composition. *Chromosome Res* *24*, 309-323.
180. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* *410*, 116-120.
181. Lahouassa, H., Daddacha, W., Hofmann, H., Ayinde, D., Logue, E.C., Dragin, L., Bloch, N., Maudet, C., Bertrand, M., Gramberg, T., *et al.* (2012). SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nature immunology* *13*, 223-228.
182. Lam, D.D., de Souza, F.S., Nasif, S., Yamashita, M., Lopez-Leal, R., Otero-Corchon, V., Meece, K., Sampath, H., Mercer, A.J., Wardlaw, S.L., *et al.* (2015). Partially redundant enhancers cooperatively maintain Mammalian pomc expression above a critical functional threshold. *PLoS Genet* *11*, e1004935.
183. Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Kochert, K., Bouhrel, M.A., Richter, J., Soler, E., *et al.* (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat Med* *16*, 571-579, 571p following 579.
184. Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* *409*, 860-921.
185. Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* *35*, 88-93.
186. Lavie, L., Medstrand, P., Schempp, W., Meese, E., and Mayer, J. (2004). Human endogenous retrovirus family HERV-K(HML-5): status, evolution, and reconstruction of an ancient betaretrovirus in the human genome. *J Virol* *78*, 8788-8798.
187. Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* *11*, 204-220.
188. Lechner, M.S., Begg, G.E., Speicher, D.W., and Rauscher, F.J. (2000). Molecular Determinants for Targeting Heterochromatin Protein 1-Mediated Gene Silencing: Direct Chromoshadow Domain-KAP-1 Corepressor Interaction Is Essential. In *Mol Cell Biol*, pp. 6449-6465.

189. Lee, A., Nolan, A., Watson, J., and Tristem, M. (2013a). Identification of an ancient endogenous retrovirus, predating the divergence of the placental mammals. *Philos Trans R Soc Lond B Biol Sci* 368, 20120503.
190. Lee, V.M., and Andrews, P.W. (1986). Differentiation of NTERA-2 clonal human embryonal carcinoma cells into neurons involves the induction of all three neurofilament proteins. *J Neurosci* 6, 514-521.
191. Lee, Y.J., Jeong, B.H., Choi, E.K., and Kim, Y.S. (2013b). Involvement of Endogenous Retroviruses in Prion Diseases. In *Pathogens*, pp. 533-543.
192. Lee, Y.N., and Bieniasz, P.D. (2007). Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog* 3, e10.
193. Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., and Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev* 24, 265-276.
194. Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* 13, 1192-1200.
195. Lester, S.N., and Li, K. (2014). Toll-like receptors in antiviral innate immunity. *J Mol Biol* 426, 1246-1264.
196. Leung, D., Du, T., Wagner, U., Xie, W., Lee, A.Y., Goyal, P., Li, Y., Szulwach, K.E., Jin, P., Lorincz, M.C., *et al.* (2014). Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proc Natl Acad Sci U S A* 111, 6690-6695.
197. Leung, D.C., Dong, K.B., Maksakova, I.A., Goyal, P., Appanah, R., Lee, S., Tachibana, M., Shinkai, Y., Lehnertz, B., Mager, D.L., *et al.* (2011). Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing. *Proc Natl Acad Sci U S A* 108, 5718-5723.
198. Li, J., Li, Z., Ruan, J., Xu, C., Tong, Y., Pan, P.W., Tempel, W., Crombet, L., Min, J., and Zang, J. (2011). Structural basis for specific binding of human MPP8 chromodomain to histone H3 methylated at lysine 9. *PLoS One* 6, e25104.
199. Li, J.M., Fan, W.S., Horsfall, A.C., Anderson, A.C., Rigby, S., Larsson, E., and Venables, P.J. (1996). The expression of human endogenous retrovirus-3 in fetal cardiac tissue and antibodies in congenital heart block. *Clin Exp Immunol* 104, 388-393.
200. Li, W., Lee, M.H., Henderson, L., Tyagi, R., Bachani, M., Steiner, J., Campanac, E., Hoffman, D.A., von Geldern, G., Johnson, K., *et al.* (2015). Human endogenous retrovirus-K contributes to motor neuron disease. *Sci Transl Med* 7, 307-153.
201. Liang, Q., Xu, Z., Xu, R., Wu, L., and Zheng, S. (2012). Expression patterns of non-coding spliced transcripts from human endogenous retrovirus HERV-H elements in colon cancer. *PLoS One* 7, e29950.
202. Liddicoat, B.J., Piskol, R., Chalk, A.M., Ramaswami, G., Higuchi, M., Hartner, J.C., Li, J.B., Seeburg, P.H., and Walkley, C.R. (2015). RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* 349, 1115-1120.

203. Ling, J., Pi, W., Bollag, R., Zeng, S., Keskinetepe, M., Saliman, H., Krantz, S., Whitney, B., and Tuan, D. (2002). The solitary long terminal repeats of ERV-9 endogenous retrovirus are conserved during primate evolution and possess enhancer activities in embryonic and hematopoietic cells. *J Virol* 76, 2410-2423.
204. Liu, R.Z., Graham, K., Glubrecht, D.D., Lai, R., Mackey, J.R., and Godbout, R. (2012). A fatty acid-binding protein 7/RXRbeta pathway enhances survival and proliferation in triple-negative breast cancer. *J Pathol* 228, 310-321.
205. Liu, X., Gao, Q., Li, P., Zhao, Q., Zhang, J., Li, J., Koseki, H., and Wong, J. (2013). UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat Commun* 4, 1563.
206. Lo, W.L., Felix, N.J., Walters, J.J., Rohrs, H., Gross, M.L., and Allen, P.M. (2009). An endogenous peptide positively selects and augments the activation and survival of peripheral CD4⁺ T cells. *Nat Immunol* 10, 1155-1161.
207. Lock, F.E., Rebollo, R., Miceli-Royer, K., Gagnier, L., Kuah, S., Babaian, A., Sistiaga-Poveda, M., Lai, C.B., Nemirovsky, O., Serrano, I., *et al.* (2014). Distinct isoform of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma. *Proc Natl Acad Sci U S A* 111, E3534-3543.
208. Looman, C., Abrink, M., Mark, C., and Hellman, L. (2002). KRAB zinc finger proteins: an analysis of the molecular mechanisms governing their increase in numbers and complexity during evolution. *Mol Biol Evol* 19, 2118-2130.
209. Lorincz, M.C., Schübeler, D., Goeke, S.C., Walters, M., Groudine, M., and Martin, D.I.K. (2000). Dynamic Analysis of Proviral Induction and De Novo Methylation: Implications for a Histone Deacetylase-Independent, Methylation Density-Dependent Mechanism of Transcriptional Repression. In *Mol Cell Biol*, pp. 842-850.
210. Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014a). Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev* 28, 2103-2119.
211. Lu, X., Sachs, F., Ramsay, L., Jacques, P.E., Goke, J., Bourque, G., and Ng, H.H. (2014b). The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol* 21, 423-425.
212. Luan, D.D., Korman, M.H., Jakubczak, J.L., and Eickbush, T.H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595-605.
213. Lukic, S., Nicolas, J.C., and Levine, A.J. (2014). The diversity of zinc-finger genes on human chromosome 19 provides an evolutionary mechanism for defense against inherited endogenous retroviruses. *Cell Death Differ* 21, 381-387.
214. Lynch, C.A., Tycko, B., Bestor, T.H., and Walsh, C.P. (2002). Reactivation of a silenced H19 gene in human rhabdomyosarcoma by demethylation of DNA but not by histone hyperacetylation. In *Mol Cancer*, p. 2.

215. Macfarlan, T.S., Gifford, W.D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., Andrews, S.E., Franco, L., Rosenfeld, M.G., Ren, B., *et al.* (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev* 25, 594-607.
216. Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57-63.
217. Magiorkinis, G., Gifford, R.J., Katzourakis, A., De Ranter, J., and Belshaw, R. (2012). Env-less endogenous retroviruses are genomic superspreaders. *Proc Natl Acad Sci U S A* 109, 7385-7390.
218. Maksakova, I.A., Thompson, P.J., Goyal, P., Jones, S.J., Singh, P.B., Karimi, M.M., and Lorincz, M.C. (2013). Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics & Chromatin* 6, 15.
219. Malik, H.S., Henikoff, S., and Eickbush, T.H. (2000). Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res* 10, 1307-1318.
220. Mangeney, M., de Parseval, N., Thomas, G., and Heidmann, T. (2001). The full-length envelope of an HERV-H human endogenous retrovirus has immunosuppressive properties. *J Gen Virol* 82, 2515-2518.
221. Mangeney, M., Renard, M., Schlecht-Louf, G., Bouallaga, I., Heidmann, O., Letzelter, C., Richaud, A., Ducos, B., and Heidmann, T. (2007). Placental syncytins: Genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. *Proc Natl Acad Sci U S A* 104, 20534-20539.
222. Manghera, M., and Douville, R.N. (2013). Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? *Retrovirology* 10, 16.
223. Marguerat, S., Wang, W.Y., Todd, J.A., and Conrad, B. (2004). Association of human endogenous retrovirus K-18 polymorphisms with type 1 diabetes. *Diabetes* 53, 852-854.
224. Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *Embo j* 24, 800-812.
225. Martin, S.L. (1991). Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol Cell Biol* 11, 4804-4807.
226. Mathias, S.L., Scott, A.F., Kazazian, H.H., Jr., Boeke, J.D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. *Science* 254, 1808-1810.
227. Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* 464, 927-931.
228. McClintock, B. (1956). Intranuclear systems controlling gene action and mutation. *Brookhaven Symp Biol*, 58-74.

229. McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E.R., Wilson, R.K., *et al.* (2001). A physical map of the human genome. *Nature* 409, 934-941.
230. Medstrand, P., van de Lagemaat, L.N., and Mager, D.L. (2002). Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res* 12, 1483-1495.
231. Meisler, M.H., and Ting, C.N. (1993). The remarkable evolutionary history of the human amylase genes. *Crit Rev Oral Biol Med* 4, 503-509.
232. Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al.* (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766-770.
233. Mi, S., Lee, X., Li, X., Veldman, G.M., Finnerty, H., Racie, L., LaVallie, E., Tang, X.Y., Edouard, P., Howes, S., *et al.* (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403, 785-789.
234. Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553-560.
235. Moran, J.V., DeBerardinis, R.J., and Kazazian, H.H., Jr. (1999). Exon shuffling by L1 retrotransposition. *Science* 283, 1530-1534.
236. Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D., and Kazazian, H.H., Jr. (1996). High frequency retrotransposition in cultured mammalian cells. *Cell* 87, 917-927.
237. Morin, R.D., Mendez-Lago, M., Mungall, A.J., Goya, R., Mungall, K.L., Corbett, R.D., Johnson, N.A., Severson, T.M., Chiu, R., Field, M., *et al.* (2011). Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 476, 298-303.
238. Morozov, V.A., Dao Thi, V.L., and Denner, J. (2013). The transmembrane protein of the human endogenous retrovirus-K (HERV-K) modulates cytokine release and gene expression. *PLoS One* 8, e70399.
239. Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903-910.
240. Murphy, V.J., Harrison, L.C., Rudert, W.A., Luppi, P., Trucco, M., Fierabracci, A., Biro, P.A., and Bottazzo, G.F. (1998). Retroviral superantigens and type 1 diabetes mellitus. *Cell* 95, 9-11.
241. Myers, L., Joedicke, J.J., Carmody, A.B., Messer, R.J., Kassiotis, G., Dudley, J.P., Dittmer, U., and Hasenkrug, K.J. (2013). IL-2-independent and TNF-alpha-dependent expansion of Vbeta5+ natural regulatory T cells during retrovirus infection. *J Immunol* 190, 5485-5495.
242. Nagy, G., Ward, J., Mosser, D.D., Koncz, A., Gergely, P., Jr., Stancato, C., Qian, Y., Fernandez, D., Niland, B., Grossman, C.E., *et al.* (2006). Regulation of CD4 expression via recycling by HRES-1/RAB4 controls susceptibility to HIV infection. *J Biol Chem* 281, 34574-34591.
243. Naito, T., Ogasawara, H., Kaneko, H., Hishikawa, T., Sekigawa, I., Hashimoto, H., and Maruyama, N. (2003). Immune abnormalities induced by human endogenous retroviral peptides: with reference to the

- pathogenesis of systemic lupus erythematosus. *J Clin Immunol* 23, 371-376.
244. Najafabadi, H.S., Mnaimneh, S., Schmitges, F.W., Garton, M., Lam, K.N., Yang, A., Albu, M., Weirauch, M.T., Radovani, E., Kim, P.M., *et al.* (2015). C2H2 zinc finger proteins greatly expand the human regulatory lexicon. *Nat Biotechnol* 33, 555-562.
 245. Nan, X., Ng, H.H., Johnson, C.A., Laherty, C.D., Turner, B.M., Eisenman, R.N., and Bird, A. (1998). Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393, 386-389.
 246. Nelson, P.N., Lever, A.M., Smith, S., Pitman, R., Murray, P., Perera, S.A., Westwood, O.M., Hay, F.C., Ejtehadi, H.D., and Booth, J.C. (1999). Molecular investigations implicate human endogenous retroviruses as mediators of anti-retroviral antibodies in autoimmune rheumatic disease. *Immunol Invest* 28, 277-289.
 247. Nielsen, A.L., Ortiz, J.A., You, J., Oulad-Abdelghani, M., Khechumian, R., Gansmuller, A., Chambon, P., and Losson, R. (1999). Interaction with members of the heterochromatin protein 1 (HP1) family and histone deacetylation are differentially involved in transcriptional silencing by members of the TIF1 family. *Embo j* 18, 6385-6395.
 248. Nigumann, P., Redik, K., Matlik, K., and Speek, M. (2002). Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics* 79, 628-634.
 249. O'Geen, H., Squazzo, S.L., Iyengar, S., Blahnik, K., Rinn, J.L., Chang, H.Y., Green, R., and Farnham, P.J. (2007). Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. *PLoS Genet* 3, e89.
 250. O'Neill, L.A., Golenbock, D., and Bowie, A.G. (2013). The history of Toll-like receptors - redefining innate immunity. In *Nat Rev Immunol* (England), pp. 453-460.
 251. Ohnuki, M., Tanabe, K., Sutou, K., Teramoto, I., Sawamura, Y., Narita, M., Nakamura, M., Tokunaga, Y., Watanabe, A., Yamanaka, S., *et al.* (2014). Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc Natl Acad Sci U S A* 111, 12426-12431.
 252. Okada, N. (1991). SINEs. *Curr Opin Genet Dev* 1, 498-504.
 253. Okada, N., Hamada, M., Ogiwara, I., and Ohshima, K. (1997). SINEs and LINEs share common 3' sequences: a review. *Gene* 205, 229-243.
 254. Okano, M., Xie, S., and Li, E. (1998). Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res* 26, 2536-2540.
 255. Oluwole, S.O., Yao, Y., Conradi, S., Kristensson, K., and Karlsson, H. (2007). Elevated levels of transcripts encoding a human retroviral envelope protein (syncytin) in muscles from patients with motor neuron disease. *Amyotroph Lateral Scler* 8, 67-72.
 256. Ono, M., Kawakami, M., and Takezawa, T. (1987). A novel human nonviral retroposon derived from an endogenous retrovirus. *Nucleic Acids Res* 15, 8725-8737.
 257. Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D., *et al.* (2007). DNMT3L connects

- unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448, 714-717.
258. Orgel, L.E., and Crick, F.H. (1980). Selfish DNA: the ultimate parasite. *Nature* 284, 604-607.
 259. Ostertag, E.M., Goodier, J.L., Zhang, Y., and Kazazian Jr, H.H. (2003). SVA Elements Are Nonautonomous Retrotransposons that Cause Disease in Humans. In *Am J Hum Genet*, pp. 1444-1451.
 260. Ostertag, E.M., and Kazazian, H.H., Jr. (2001). Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res* 11, 2059-2065.
 261. Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* 10, 475-478.
 262. Paludan, S.R., and Bowie, A.G. (2013). Immune sensing of DNA. *Immunity* 38, 870-880.
 263. Parkin, J., and Cohen, B. (2001). An overview of the immune system. *Lancet* 357, 1777-1789.
 264. Pascale, E., Liu, C., Valle, E., Usdin, K., and Furano, A.V. (1993). The evolution of long interspersed repeated DNA (L1, LINE 1) as revealed by the analysis of an ancient rodent L1 DNA family. *J Mol Evol* 36, 9-20.
 265. Payne, L.N., and Chubb, R.C. (1968). Studies on the nature and genetic control of an antigen in normal chick embryos which reacts in the COFAL test. *J Gen Virol* 3, 379-391.
 266. Peaston, A.E., Evsikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D., and Knowles, B.B. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev Cell* 7, 597-606.
 267. Peng, H., Begg, G.E., Schultz, D.C., Friedman, J.R., Jensen, D.E., Speicher, D.W., and Rauscher, F.J., 3rd (2000). Reconstitution of the KRAB-KAP-1 repressor complex: a model system for defining the molecular anatomy of RING-B box-coiled-coil domain-mediated protein-protein interactions. *J Mol Biol* 295, 1139-1162.
 268. Perl, A. (2003). Role of endogenous retroviruses in autoimmune diseases. *Rheum Dis Clin North Am* 29, 123-143.
 269. Perl, A., Colombo, E., Dai, H., Agarwal, R., Mark, K.A., Banki, K., Poiesz, B.J., Phillips, P.E., Hoch, S.O., Reveille, J.D., *et al.* (1995). Antibody reactivity to the HRES-1 endogenous retroviral element identifies a subset of patients with systemic lupus erythematosus and overlap syndromes. Correlation with antinuclear antibodies and HLA class II alleles. *Arthritis Rheum* 38, 1660-1671.
 270. Perron, H., Germi, R., Bernard, C., Garcia-Montojo, M., Deluen, C., Farinelli, L., Faucard, R., Veas, F., Stefas, I., Fabriek, B.O., *et al.* (2012). Human endogenous retrovirus type W envelope expression in blood and brain cells provides new insights into multiple sclerosis disease. In *Mult Scler*, pp. 1721-1736.
 271. Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., *et al.* (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* 107, 323-337.

272. Phillips, J.E., Thyagarajan, B., and Calos, M.P. (1999). Epstein-Barr virus plasmid model system for analyzing recombination in human cells. *Plasmid* 41, 198-206.
273. Pi, W., Zhu, X., Wu, M., Wang, Y., Fulzele, S., Eroglu, A., Ling, J., and Tuan, D. (2010). Long-range function of an intergenic retrotransposon. *Proc Natl Acad Sci U S A* 107, 12992-12997.
274. Pichlmair, A., Schulz, O., Tan, C.P., Rehwinkel, J., Kato, H., Takeuchi, O., Akira, S., Way, M., Schiavo, G., and Reis e Sousa, C. (2009). Activation of MDA5 Requires Higher-Order RNA Structures Generated during Virus Infection. In *J Virol*, pp. 10761-10769.
275. Pikó, L., Hammons, M.D., and Taylor, K.D. (1984). Amounts, synthesis, and some properties of intracisternal A particle-related RNA in early mouse embryos. *Proc Natl Acad Sci U S A* 81, 488-492.
276. Plank, J.L., and Dean, A. (2014). Enhancer function: mechanistic and genome-wide insights come together. *Mol Cell* 55, 5-14.
277. Poznanski, A.A., and Calarco, P.G. (1991). The expression of intracisternal A particle genes in the preimplantation mouse embryo. *Dev Biol* 143, 271-281.
278. Punkosdy, G.A., Blain, M., Glass, D.D., Lozano, M.M., O'Mara, L., Dudley, J.P., Ahmed, R., and Shevach, E.M. (2011). Regulatory T-cell expansion during chronic viral infection is dependent on endogenous retroviral superantigens. *Proc Natl Acad Sci U S A* 108, 3677-3682.
279. Pérot, P., Mullins, C.S., Naville, M., Bressan, C., Hühns, M., Gock, M., Kühn, F., Volff, J.N., Trillet-Lenoir, V., Linnebacher, M., *et al.* (2015). Expression of young HERV-H loci in the course of colorectal carcinoma and correlation with molecular subtypes. *Oncotarget* 6, 40095-40111.
280. Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279-283.
281. Rambaud, J., Desroches, J., Balsalobre, A., and Drouin, J. (2009). TIF1beta/KAP-1 is a coactivator of the orphan nuclear receptor NGFI-B/Nur77. *J Biol Chem* 284, 14147-14156.
282. Rangel, L.B., Sherman-Baust, C.A., Wernyj, R.P., Schwartz, D.R., Cho, K.R., and Morin, P.J. (2003). Characterization of novel human ovarian cancer-specific transcripts (HOSTs) identified by serial analysis of gene expression. *Oncogene* 22, 7225-7232.
283. Rebollo, R., Karimi, M.M., Bilenky, M., Gagnier, L., Miceli-Royer, K., Zhang, Y., Goyal, P., Keane, T.M., Jones, S., Hirst, M., *et al.* (2011). Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms. *PLoS Genet* 7, e1002301.
284. Rebollo, R., Romanish, M.T., and Mager, D.L. (2012). Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet* 46, 21-42.
285. Reddy, K.L., Zullo, J.M., Bertolino, E., and Singh, H. (2008). Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* 452, 243-247.
286. Reichmann, J., Crichton, J.H., Madej, M.J., Taggart, M., Gautier, P., Garcia-Perez, J.L., Meehan, R.R., and Adams, I.R. (2012). Microarray analysis of LTR retrotransposon silencing identifies Hdac1 as a regulator of

- retrotransposon expression in mouse embryonic stem cells. *PLoS Comput Biol* 8, e1002486.
287. Reik, W. (2007). Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* 447, 425-432.
 288. Reilly, M.T., Faulkner, G.J., Dubnau, J., Ponomarev, I., and Gage, F.H. (2013). The Role of Transposable Elements in Health and Diseases of the Central Nervous System. In *J Neurosci*, pp. 17577-17586.
 289. Renaudineau, Y., Hillion, S., Saraux, A., Mageed, R.A., and Youinou, P. (2005). An alternative exon 1 of the CD5 gene regulates CD5 expression in human B lymphocytes. *Blood* 106, 2781-2789.
 290. Reynier, F., Verjat, T., Turrel, F., Imbert, P.E., Marotte, H., Mouglin, B., and Miossec, P. (2009). Increase in human endogenous retrovirus HERV-K (HML-2) viral load in active rheumatoid arthritis. *Scand J Immunol* 70, 295-299.
 291. Ribet, D., Harper, F., Dupressoir, A., Dewannieux, M., Pierron, G., and Heidmann, T. (2008). An infectious progenitor for the murine IAP retrotransposon: emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Res* 18, 597-609.
 292. Rigby, R.E., Webb, L.M., Mackenzie, K.J., Li, Y., Leitch, A., Reijns, M.A., Lundie, R.J., Revuelta, A., Davidson, D.J., Diebold, S., *et al.* (2014). RNA:DNA hybrids are a novel molecular pattern sensed by TLR9. *Embo j* 33, 542-558.
 293. Robinson, H.L., Astrin, S.M., Senior, A.M., and Salazar, F.H. (1981). Host Susceptibility to endogenous viruses: defective, glycoprotein-expressing proviruses interfere with infections. *J Virol* 40, 745-751.
 294. Rolland, A., Jouvin-Marche, E., Saresella, M., Ferrante, P., Cavaretta, R., Creange, A., Marche, P., and Perron, H. (2005). Correlation between disease severity and in vitro cytokine production mediated by MSRV (multiple sclerosis associated retroviral element) envelope protein in patients with multiple sclerosis. *J Neuroimmunol* 160, 195-203.
 295. Rolland, A., Jouvin-Marche, E., Viret, C., Faure, M., Perron, H., and Marche, P.N. (2006). The envelope protein of a human endogenous retrovirus-W family activates innate immunity through CD14/TLR4 and promotes Th1-like responses. *J Immunol* 176, 7636-7644.
 296. Romanish, M.T., Lock, W.M., van de Lagemaat, L.N., Dunn, C.A., and Mager, D.L. (2007). Repeated Recruitment of LTR Retrotransposons as Promoters by the Anti-Apoptotic Locus NAIP during Mammalian Evolution. *PLoS Genet* 3, e10.
 297. Roulois, D., Loo Yau, H., Singhania, R., Wang, Y., Danesh, A., Shen, S.Y., Han, H., Liang, G., Jones, P.A., Pugh, T.J., *et al.* (2015). DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* 162, 961-973.
 298. Rowe, H.M., Friedli, M., Offner, S., Verp, S., Mesnard, D., Marquis, J., Aktas, T., and Trono, D. (2013a). De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development (Cambridge, England)* 140, 519-529.
 299. Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., *et al.* (2010).

KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* 463, 237-240.

300. Rowe, H.M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T.S., Tarabay, Y., Viville, S., Jakobsson, J., Pfaff, S.L., and Trono, D. (2013b). TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res* 23, 452-461.
301. Ryoo, J., Choi, J., Oh, C., Kim, S., Seo, M., Kim, S.Y., Seo, D., Kim, J., White, T.E., Brandariz-Nunez, A., *et al.* (2014). The ribonuclease activity of SAMHD1 is required for HIV-1 restriction. *Nat Med* 20, 936-941.
302. Sacha, J.B., Kim, I.J., Chen, L., Ullah, J.H., Goodwin, D.A., Simmons, H.A., Schenkman, D.I., von Pelchrzim, F., Gifford, R.J., Nimityongskul, F.A., *et al.* (2012). Vaccination with cancer- and HIV infection-associated endogenous retrotransposable elements is safe and immunogenic. *J Immunol* 189, 1467-1479.
303. Sadic, D., Schmidt, K., Groh, S., Kondofersky, I., Ellwart, J., Fuchs, C., Theis, F.J., and Schotta, G. (2015). Atrx promotes heterochromatin formation at retrotransposons. *EMBO Rep* 16, 836-850.
304. Santoni de Sio, F., Barde, I., Offner, S., Kapopoulou, A., Genolet, R., Corsinotti, A., Bojkowska, K., Thomas, J., Luescher, I., Pinschewer, D., *et al.* (2012a). KAP1 Regulates Gene Networks Controlling T cell Development and Responsiveness. *FASEB J* 26, 4561-4575.
305. Santoni de Sio, F.R., Massacand, J., Barde, I., Offner, S., Corsinotti, A., Kapopoulou, A., Bojkowska, K., Dagklis, A., Fernandez, M., Ghia, P., *et al.* (2012b). KAP1 regulates gene networks controlling mouse B-lymphoid cell differentiation and function. *Blood* 119, 4675-4685.
306. Santoni, F.A., Guerra, J., and Luban, J. (2012). HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology* 9, 111.
307. Satou, A., Taira, T., Iguchi-Ariga, S.M., and Ariga, H. (2001). A novel transrepression pathway of c-Myc. Recruitment of a transcriptional corepressor complex to c-Myc by MM-1, a c-Myc-binding protein. *J Biol Chem* 276, 46562-46567.
308. Scarfo, I., Pellegrino, E., Mereu, E., Kwee, I., Agnelli, L., Bergaggio, E., Garaffo, G., Vitale, N., Caputo, M., Machiorlatti, R., *et al.* (2016). Identification of a new subclass of ALK-negative ALCL expressing aberrant levels of ERBB4 transcripts. *Blood* 127, 221-232.
309. Scherer, W.F., Syverton, J.T., and Gey, G.O. (1953). STUDIES ON THE PROPAGATION IN VITRO OF POLIOMYELITIS VIRUSES : IV. VIRAL MULTIPLICATION IN A STABLE STRAIN OF HUMAN MALIGNANT EPITHELIAL CELLS (STRAIN HELA) DERIVED FROM AN EPIDERMOID CARCINOMA OF THE CERVIX. *J Exp Med* 97, 695-710.
310. Schlesinger, S., Lee, A., Wang, G.Z., Green, L., and Goff, S.P. (2013). Proviral Silencing in Embryonic Cells is Regulated by Yin Yang 1. *Cell Rep* 4, 50-58.
311. Schmitges, F.W., Radovani, E., Najafabadi, H.S., Barazandeh, M., Campitelli, L.F., Yin, Y., Jolma, A., Zhong, G., Guo, H., Kanagalingam, T., *et al.* (2016). Multiparameter functional diversity of human C2H2 zinc finger proteins. *Genome Res* 26, 1742-1752.

312. Schmitt, K., Reichrath, J., Roesch, A., Meese, E., and Mayer, J. (2013). Transcriptional profiling of human endogenous retrovirus group HERV-K(HML-2) loci in melanoma. *Genome Biol Evol* 5, 307-328.
313. Schoggins, J.W., MacDuff, D.A., Imanaka, N., Gainey, M.D., Shrestha, B., Eitson, J.L., Mar, K.B., Richardson, R.B., Ratushny, A.V., Litvak, V., *et al.* (2014). Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. *Nature* 505, 691-695.
314. Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev* 16, 919-932.
315. Schultz, D.C., Friedman, J.R., and Rauscher, F.J., 3rd (2001). Targeting histone deacetylase complexes via KRAB-zinc finger proteins: the PHD and bromodomains of KAP-1 form a cooperative unit that recruits a novel isoform of the Mi-2alpha subunit of NuRD. *Genes Dev* 15, 428-443.
316. Schübeler, D., Lorincz, M.C., Cimbara, D.M., Telling, A., Feng, Y.Q., Bouhassira, E.E., and Groudine, M. (2000). Genomic Targeting of Methylated DNA: Influence of Methylation on Transcription, Replication, Chromatin Structure, and Histone Acetylation. In *Mol Cell Biol*, pp. 9103-9112.
317. Shaikh, T.H., Roy, A.M., Kim, J., Batzer, M.A., and Deininger, P.L. (1997). cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts. *J Mol Biol* 271, 222-234.
318. Sharma, P., Azebi, S., England, P., Christensen, T., Moller-Larsen, A., Petersen, T., Batsche, E., and Muchardt, C. (2012). Citrullination of histone H3 interferes with HP1-mediated transcriptional repression. *PLoS Genet* 8, e1002934.
319. Sharma, S., DeOliveira, R.B., Kalantari, P., Parroche, P., Goutagny, N., Jiang, Z., Chan, J., Bartholomeu, D.C., Lauw, F., Hall, J.P., *et al.* (2011). Innate immune recognition of an AT-rich stem-loop DNA motif in the *Plasmodium falciparum* genome. *Immunity* 35, 194-207.
320. Sheehy, A.M., Gaddis, N.C., Choi, J.D., and Malim, M.H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 418, 646-650.
321. Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119, 941-953.
322. Sicat, J., Sutkowski, N., and Huber, B.T. (2005). Expression of human endogenous retrovirus HERV-K18 superantigen is elevated in juvenile rheumatoid arthritis. *J Rheumatol* 32, 1821-1831.
323. Singer, M.F. (1982). Highly repeated sequences in mammalian genomes. *Int Rev Cytol* 76, 67-112.
324. Singh, K., Cassano, M., Planet, E., Sebastian, S., Jang, S.M., Sohi, G., Faralli, H., Choi, J., Youn, H.D., Dilworth, F.J., *et al.* (2015). A KAP1 phosphorylation switch controls MyoD function during skeletal muscle differentiation. *Genes Dev* 29, 513-525.

325. Siomi, M.C., Sato, K., Pezic, D., and Aravin, A.A. (2011). PIWI-interacting small RNAs: the vanguard of genome defence. *Nat Rev Mol Cell Biol* 12, 246-258.
326. Smit, A.F., Toth, G., Riggs, A.D., and Jurka, J. (1995). Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J Mol Biol* 246, 401-417.
327. Smothers, J.F., and Henikoff, S. (2000). The HP1 chromo shadow domain binds a consensus peptide pentamer. *Curr Biol* 10, 27-30.
328. Sokol, M., Jessen, K.M., and Pedersen, F.S. (2015). Human endogenous retroviruses sustain complex and cooperative regulation of gene-containing loci and unannotated megabase-sized regions. *Retrovirology* 12, 32.
329. Sommerfelt, M.A., and Weiss, R.A. (1990). Receptor interference groups of 20 retroviruses plating on human cells. *Virology* 176, 58-69.
330. Song, S.U., Gerasimova, T., Kurkulos, M., Boeke, J.D., and Corces, V.G. (1994). An env-like protein encoded by a Drosophila retroelement: evidence that gypsy is an infectious retrovirus. *Genes Dev* 8, 2046-2057.
331. Speek, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol Cell Biol* 21, 1973-1985.
332. Sripathy, S.P., Stevens, J., and Schultz, D.C. (2006). The KAP1 corepressor functions to coordinate the assembly of de novo HP1-demarcated microenvironments of heterochromatin required for KRAB zinc finger protein-mediated transcriptional repression. *Molecular and cellular biology* 26, 8623-8638.
333. Stetson, D.B., Ko, J.S., Heidmann, T., and Medzhitov, R. (2008). Trex1 Prevents Cell-Intrinsic Initiation of Autoimmunity. *Cell* 134, 587-598.
334. Suetake, I., Shinozaki, F., Miyagawa, J., Takeshima, H., and Tajima, S. (2004). DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J Biol Chem* 279, 27816-27823.
335. Sugimoto, J., Sugimoto, M., Bernstein, H., Jinno, Y., and Schust, D. (2013). A novel human endogenous retroviral protein inhibits cell-cell fusion. *Sci Rep* 3, 1462.
336. Sun, L., Wu, J., Du, F., Chen, X., and Chen, Z.J. (2013). Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway. *Science (New York, NY)* 339, 786-791.
337. Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M.P., and Wang, T. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res* 24, 1963-1976.
338. Sundquist, W.I., and Kräusslich, H.G. (2012). HIV-1 Assembly, Budding, and Maturation. In *Cold Spring Harb Perspect Med*, p. a006924.
339. Suntsova, M., Gogvadze, E.V., Salozhin, S., Gaifullin, N., Eroshkin, F., Dmitriev, S.E., Martynova, N., Kulikov, K., Malakhova, G., Tukhbatova, G., *et al.* (2013). Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene PRODH. *Proc Natl Acad Sci U S A* 110, 19472-19477.
340. Svensson, K., Mattsson, R., James, T.C., Wentzel, P., Pilartz, M., MacLaughlin, J., Miller, S.J., Olsson, T., Eriksson, U.J., and Ohlsson, R. (1998). The paternal allele of the H19 gene is progressively silenced

during early mouse development: the acetylation status of histones may be involved in the generation of variegated expression patterns. *Development* 125, 61-69.

341. Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G.J., and Schultz, R.M. (2004). RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev Biol* 269, 276-285.
342. Swergold, G.D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol Cell Biol* 10, 6718-6729.
343. Switzer, W.M., Salemi, M., Shanmugam, V., Gao, F., Cong, M.E., Kuiken, C., Bhullar, V., Beer, B.E., Vallet, D., Gautier-Hion, A., *et al.* (2005). Ancient co-speciation of simian foamy viruses and primates. *Nature* 434, 376-380.
344. Szak, S.T., Pickeral, O.K., Makalowski, W., Boguski, M.S., Landsman, D., and Boeke, J.D. (2002). Molecular archeology of L1 insertions in the human genome. *Genome Biol* 3, 1-18.
345. Szondy, Z., Garabuczi, E., Joos, G., Tsay, G.J., and Sarang, Z. (2014). Impaired clearance of apoptotic cells in chronic inflammatory diseases: therapeutic implications. *Front Immunol* 5, 354.
346. Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H., and Shinkai, Y. (2008). G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. In *EMBO J*, pp. 2681-2690.
347. Takaoka, A., Wang, Z., Choi, M.K., Yanai, H., Negishi, H., Ban, T., Lu, Y., Miyagishi, M., Kodama, T., Honda, K., *et al.* (2007). DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* 448, 501-505.
348. Tang, D., Kang, R., Coyne, C.B., Zeh, H.J., and Lotze, M.T. (2012). PAMPs and DAMPs: Signal 0s that Spur Autophagy and Immunity. *Immunol Rev* 249, 158-175.
349. Tchasovnikarova, I.A., Timms, R.T., Douse, C.H., Roberts, R.C., Dougan, G., Kingston, R.E., Modis, Y., and Lehner, P.J. (2017). Hyperactivation of HUSH complex function by Charcot-Marie-Tooth disease mutation in MORC2. *Nat Genet* 49, 1035-1044.
350. Tchasovnikarova, I.A., Timms, R.T., Matheson, N.J., Wals, K., Antrobus, R., Göttgens, B., Dougan, G., Dawson, M.A., and Lehner, P.J. (2015). Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science* 348, 1481-1485.
351. Teich, N.M., Weiss, R.A., Martin, G.R., and Lowy, D.R. (1977). Virus infection of murine teratocarcinoma stem cell lines. *Cell* 12, 973-982.
352. Thomas, J.H., and Schneider, S. (2011). Coevolution of retroelements and tandem zinc finger genes. *Genome Res* 21, 1800-1812.
353. Thompson, P.J., Dulberg, V., Moon, K.M., Foster, L.J., Chen, C., Karimi, M.M., and Lorincz, M.C. (2015). hnRNP K coordinates transcriptional silencing by SETDB1 in embryonic stem cells. *PLoS Genet* 11, e1004933.
354. Thompson, P.J., Macfarlan, T.S., and Lorincz, M.C. (2016). Long Terminal Repeats: From Parasitic Elements to Building Blocks of the Transcriptional Regulatory Repertoire. *Mol Cell* 62, 766-776.
355. Tomlins, S.A., Laxman, B., Dhanasekaran, S.M., Helgeson, B.E., Cao, X., Morris, D.S., Menon, A., Jing, X., Cao, Q., Han, B., *et al.* (2007). Distinct

classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448**, 595-599.

356. Tugnet, N., Rylance, P., Roden, D., Trela, M., and Nelson, P. (2013). Human Endogenous Retroviruses (HERVs) and Autoimmune Rheumatic Disease: Is There a Link? In *Open Rheumatol J*, pp. 13-21.
357. Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., Tieng, V., Quenneville, S., and Trono, D. (2014). Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res* **24**, 1260-1270.
358. Uesaka, M., Nishimura, O., Go, Y., Nakashima, K., Agata, K., and Imamura, T. (2014). Bidirectional promoters are the major source of gene activation-associated non-coding RNAs in mammals. *BMC Genomics* **15**, 35.
359. Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., *et al.* (2010). Towards a knowledge-based Human Protein Atlas. In *Nat Biotechnol* (United States), pp. 1248-1250.
360. Ullu, E., and Tschudi, C. (1984). Alu sequences are processed 7SL RNA genes. *Nature* **312**, 171-172.
361. Underhill, C., Qutob, M.S., Yee, S.P., and Torchia, J. (2000). A novel nuclear receptor corepressor complex, N-CoR, contains components of the mammalian SWI/SNF complex and the corepressor KAP-1. *J Biol Chem* **275**, 40463-40470.
362. Unterholzner, L., Keating, S.E., Baran, M., Horan, K.A., Jensen, S.B., Sharma, S., Sirois, C.M., Jin, T., Latz, E., Xiao, T.S., *et al.* (2010). IFI16 is an innate immune sensor for intracellular DNA. *Nat Immunol* **11**, 997-1004.
363. Varela, M., Spencer, T.E., Palmarini, M., and Arnaud, F. (2009). Friendly viruses: the special relationship between endogenous retroviruses and their host. *Ann N Y Acad Sci* **1178**, 157-172.
364. Vassetzky, N.S., and Kramerov, D.A. (2013). SINEBase: a database and tool for SINE analysis. *Nucleic Acids Res* **41**, D83-89.
365. Veselovska, L., Smallwood, S.A., Saadeh, H., Stewart, K.R., Krueger, F., Maupetit-Mehouas, S., Arnaud, P., Tomizawa, S., Andrews, S., and Kelsey, G. (2015). Deep sequencing and de novo assembly of the mouse oocyte transcriptome define the contribution of transcription to the DNA methylation landscape. *Genome Biol* **16**, 209.
366. Vogel, M.J., Guelen, L., de Wit, E., Hupkes, D.P., Lodén, M., Talhout, W., Feenstra, M., Abbas, B., Classen, A.K., and van Steensel, B. (2006). Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res* **16**, 1493-1504.
367. Voliva, C.F., Martin, S.L., Hutchison, C.A., 3rd, and Edgell, M.H. (1984). Dispersal process associated with the L1 family of interspersed repetitive DNA sequences. *J Mol Biol* **178**, 795-813.
368. Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20**, 116-117.
369. Walter, M., Teissandier, A., Perez-Palacios, R., and Bourc'his, D. (2016). An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* **5**.

370. Wang, C., and Goff, S.P. (2017). Differential control of retrovirus silencing in embryonic cells by proteasomal regulation of the ZFP809 retroviral repressor. *Proc Natl Acad Sci U S A* *114*, E922-e930.
371. Wang, G.Z., and Goff, S.P. (2015). Regulation of Yin Yang 1 by Tyrosine Phosphorylation. *J Biol Chem* *290*, 21890-21900.
372. Wang, G.Z., Wolf, D., and Goff, S.P. (2014a). EBP1, a novel host factor involved in primer binding site-dependent restriction of moloney murine leukemia virus in embryonic cells. *J Virol* *88*, 1825-1829.
373. Wang, H., Xing, J., Grover, D., Hedges, D.J., Han, K., Walker, J.A., and Batzer, M.A. (2005). SVA elements: a hominid-specific retroposon family. *J Mol Biol* *354*, 994-1007.
374. Wang, J., Xie, G., Singh, M., Ghanbarian, A.T., Rasko, T., Szvetnik, A., Cai, H., Besser, D., Prigione, A., Fuchs, N.V., *et al.* (2014b). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* *516*, 405-409.
375. Wang-Johanning, F., Radvanyi, L., Rycaj, K., Plummer, J.B., Yan, P., Sastry, K.J., Piyathilake, C.J., Hunt, K.K., and Johanning, G.L. (2008). Human endogenous retrovirus K triggers an antigen-specific immune response in breast cancer patients. *Cancer Res* *68*, 5869-5877.
376. Weber, F., Wagner, V., Rasmussen, S.B., Hartmann, R., and Paludan, S.R. (2006). Double-stranded RNA is produced by positive-strand RNA viruses and DNA viruses but not in detectable amounts by negative-strand RNA viruses. *J Virol* *80*, 5059-5064.
377. Wei, W., Gilbert, N., Ooi, S.L., Lawler, J.F., Ostertag, E.M., Kazazian, H.H., Boeke, J.D., and Moran, J.V. (2001). Human L1 retrotransposition: cis preference versus trans complementation. *Mol Cell Biol* *21*, 1429-1439.
378. Weiner, A.M. (1980). An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome. *Cell* *22*, 209-218.
379. Weiner, A.M. (2002). SINEs and LINEs: the art of biting the hand that feeds you. *Curr Opin Cell Biol* *14*, 343-350.
380. Wolf, D., Cammas, F., Losson, R., and Goff, S.P. (2008a). Primer binding site-dependent restriction of murine leukemia virus requires HP1 binding by TRIM28. *J Virol* *82*, 4675-4679.
381. Wolf, D., and Goff, S.P. (2007). TRIM28 Mediates Primer Binding Site-Targeted Silencing of Murine Leukemia Virus in Embryonic Cells. *Cell* *131*, 46-57.
382. Wolf, D., and Goff, S.P. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* *458*, 1201-1204.
383. Wolf, D., Hug, K., and Goff, S.P. (2008b). TRIM28 mediates primer binding site-targeted silencing of Lys1,2 tRNA-utilizing retroviruses in embryonic cells. *Proceedings of the National Academy of Sciences of the United States of America* *105*, 12521-12526.
384. Wolf, G., Rebollo, R., Karimi, M.M., Ewing, A.D., Kamada, R., Wu, W., Wu, B., Bachu, M., Ozato, K., Faulkner, G.J., *et al.* (2017). On the role of H3.3 in retroviral silencing. *Nature* *548*, E1-e3.
385. Wolf, G., Yang, P., Fuchtbauer, A.C., Fuchtbauer, E.M., Silva, A.M., Park, C., Wu, W., Nielsen, A.L., Pedersen, F.S., and Macfarlan, T.S. (2015). The

- KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev* 29, 538-554.
386. Wu, J., Zhou, T., He, J., and Mountz, J.D. (1993). Autoimmune disease in mice due to integration of an endogenous retrovirus in an apoptosis gene. *J Exp Med* 178, 461-468.
 387. Xie, M., Hong, C., Zhang, B., Lowdon, R., Xing, X., Li, D., Zhou, X., Lee, H.J., Maire, C.L., Ligon, K.L., *et al.* (2013). DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* 45, 836-841.
 388. Xu, J., Zoltick, P.W., Gamero, A.M., and Gallucci, S. (2014). TLR ligands up-regulate Trex1 expression in murine conventional dendritic cells through type I Interferon and NF-kappaB-dependent signaling pathways. *J Leukoc Biol* 96, 93-103.
 389. Yan, N., Regalado-Magdos, A.D., Stiggelbout, B., Lee-Kirsch, M.A., and Lieberman, J. (2010). The cytosolic exonuclease TREX1 inhibits the innate immune response to human immunodeficiency virus type 1. *Nature immunology* 11, 1005-1013.
 390. Yan, Q., Huang, J., Fan, T., Zhu, H., and Muegge, K. (2003). Lsh, a modulator of CpG methylation, is crucial for normal histone methylation. In *EMBO J*, pp. 5154-5162.
 391. Yang, B.X., El Farran, C.A., Guo, H.C., Yu, T., Fang, H.T., Wang, H.F., Schlesinger, S., Seah, Y.F., Goh, G.Y., Neo, S.P., *et al.* (2015). Systematic Identification of Factors for Provirus Silencing in Embryonic Stem Cells. *Cell* 163, 230-245.
 392. Yang, N., and Kazazian, H.H., Jr. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat Struct Mol Biol* 13, 763-771.
 393. Yoder, J.A., Walsh, C.P., and Bestor, T.H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 13, 335-340.
 394. Yoshinobu, K., Baudino, L., Santiago-Raber, M.L., Morito, N., Dunand-Sauthier, I., Morley, B.J., Evans, L.H., and Izui, S. (2009). Selective up-regulation of intact, but not defective env RNAs of endogenous modified polytropic retrovirus by the Sgp3 locus of lupus-prone mice. *J Immunol* 182, 8094-8103.
 395. Yotsuyanagi, Y., and Szollosi, D. (1981). Early mouse embryo intracisternal particle: Fourth type of retrovirus-like particle associated with the mouse. *J Natl Cancer Inst* 67, 677-685.
 396. You, A., Tong, J.K., Grozinger, C.M., and Schreiber, S.L. (2001). CoREST is an integral component of the CoREST- human histone deacetylase complex. In *Proc Natl Acad Sci U S A*, pp. 1454-1458.
 397. Young, G.R., Ploquin, M.J., Eksmond, U., Wadwa, M., Stoye, J.P., and Kassiotis, G. (2012). Negative selection by an endogenous retrovirus promotes a higher-avidity CD4+ T cell response to retroviral infection. *PLoS Pathog* 8, e1002709.
 398. Yu, P., Lübben, W., Slomka, H., Gebler, J., Konert, M., Cai, C., Neubrandt, L., Prazeres da Costa, O., Paul, S., Dehnert, S., *et al.* (2012). Nucleic acid-sensing Toll-like receptors are essential for the control of endogenous retrovirus viremia and ERV-induced tumors. *Immunity* 37, 867-879.

399. Zeng, L., Yap, K.L., Ivanov, A.V., Wang, X., Mujtaba, S., Plotnikova, O., Rauscher, F.J., 3rd, and Zhou, M.M. (2008). Structural insights into human KAP1 PHD finger-bromodomain and its role in gene silencing. *Nat Struct Mol Biol* 15, 626-633.
400. Zeng, M., Hu, Z., Shi, X., Li, X., Zhan, X., Li, X.D., Wang, J., Choi, J.H., Wang, K.W., Purrington, T., *et al.* (2014). MAVS, cGAS, and endogenous retroviruses in T-independent B cell responses. *Science* 346, 1486-1492.
401. Zhao, K., Du, J., Han, X., Goodier, J.L., Li, P., Zhou, X., Wei, W., Evans, S.L., Li, L., Zhang, W., *et al.* (2013). Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutières syndrome-related SAMHD1. *Cell reports* 4, 1108-1115.
402. Zhu, H., Geiman, T.M., Xi, S., Jiang, Q., Schmidtmann, A., Chen, T., Li, E., and Muegge, K. (2006). Lsh is involved in de novo methylation of DNA. *Embo j* 25, 335-345.
403. Zuin, J., Dixon, J.R., van der Reijden, M.I., Ye, Z., Kolovos, P., Brouwer, R.W., van de Corput, M.P., van de Werken, H.J., Knoch, T.A., van, I.W.F., *et al.* (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A* 111, 996-1001.
404. Zuniga, E.I., Hahm, B., and Oldstone, M.B. (2007). Type I interferon during viral infections: multiple triggers for a multifunctional mediator. *Curr Top Microbiol Immunol* 316, 337-357.

Appendix A



Epigenetic control of retrotransposons in adult tissues: implications for immune regulation

Christopher HC Tie and Helen M Rowe

Retrotransposons tune immune reactivity in differentiated cells because when they are transcribed, their nucleic acids can be viewed as non-self leading to innate immune sensing. Most retrotransposons, however, are subject to transcriptional regulation by a multitude of epigenetic pathways, which have coevolved with them for millions of years. While a lot is known about the epigenetic control of retrotransposons in germ cells and early embryos, surprisingly little is understood about these pathways in adult tissues, particularly in human cells. Recent evidence suggests that retrotransposon repression persists in differentiated cells and is dynamic. Future insight into this topic may teach us how to reactivate or silence specific retrotransposon families, to promote anti-tumor immunity or dampen autoimmunity through epigenetic modulation.

Address

Division of Infection and Immunity, University College London, 90 Gower Street, London WC1E 6BT, UK

Corresponding author: Rowe, Helen M (h.rowe@ucl.ac.uk)

Current Opinion in Virology 2017, 25:28–33

This review comes from a themed issue on **Paleovirology**

Edited by **Aris Katzourakis**

<http://dx.doi.org/10.1016/j.coviro.2017.06.007>

1879-6257/© 2017 Published by Elsevier B.V.

Introduction

Transposable elements (TEs) are mobile genetic elements aptly named ‘controlling elements’ by Barbara McClintock in the 1950s because of their ability to control cellular genes [1]. Since their discovery, the advancement of genome sequencing and bioinformatics technologies has led to the identification of a huge number of TE families, the functions of which are now being unraveled. Retrotransposons are particularly relevant TEs because they can replicate through an RNA intermediate, allowing them to insert new DNA copies of themselves into the genome. This has enabled them to accumulate over millions of years and they now comprise more than half of the human genome [2]. Over time, deleterious insertions are negatively selected while those beneficial to the host become co-opted and fixed in the genome [3]. Co-opted retrotransposons are often from the endogenous

retrovirus (ERV) class, which dates back hundreds of millions of years [4]. Although retrotransposons can be beneficial as a driving force behind the evolution of new genes and non-coding DNA [5,6], they can also compromise genome and transcriptome integrity [7]. A multitude of epigenetic pathways, therefore, act in early development to constrain their transcription and some of these strategies remain active in differentiated cells.

Epigenetics refers to modifications on chromatin, rather than DNA sequence alterations, which lead to heritable effects on gene expression. Chromatin is subject to histone modifications and cytosine methylation and distinct epigenetic marks are associated with an active and silent gene expression state. For example, acetylation of histone 3 at the lysine residue at position 27 (H3K27ac) is a chromatin signature associated with active genes and enhancers [8], whereas trimethylation of histone 3 at the lysine at position 9 (H3K9me3) correlates with heterochromatin and gene silencing [9–11]. Retrotransposons direct both genetic and epigenetic heritable traits because they can integrate into our genome in the germ line and orchestrate epigenetic alterations through the recruitment of transcription factors to their regulatory elements [12].

Since half of the human genome is derived from retrotransposons, it can be viewed as ‘non-self’ because of its viral origin. Although most human retrotransposons are no longer mobile, expression of their nucleic acids and proteins can lead to the formation of pathogen-associated molecular patterns (PAMPs) and antigens that we refer to here as ‘neo-antigens’ that could potentially elicit an immunological response. Retrotransposons, are therefore situated at the interface of immune reactivity; when enriched in silent chromatin they are transcriptionally inactive and immune masked, whereas when expressed they may trigger innate and adaptive immunity [13^{••},14^{••},15]. In this review, we will discuss the mechanisms in place to maintain retrotransposons silent in differentiated cells and the implications of these pathways. We will focus here on chromatin readers, writers and erasers and the KAP1 and KRAB-ZNF system. The role of small RNAs, while important, is beyond the scope of this review and we direct the reader to a recent review covering this topic [16].

Epigenetic pathways constraining retrotransposons

Epigenetic silencing of retrotransposons takes place in early embryos and in differentiated tissues, epigenetic

states link back to patterns established during development [17[•]]. The identification of factors that maintain repression in postembryonic tissues and how dynamic chromatin marks are in differentiated cell types is a fascinating and emerging area of research. Moreover, existing work relies largely on mouse models so it will be crucial to establish parallels in human cells, where the precise retrotransposons, their activity and their coevolution patterns with genes are distinct [18].

Chromatin readers, writers and erasers

Chromatin-associated proteins induce epigenetic changes to either histones or DNA and mediate downstream biological functions. Promoter cytosine methylation occurs at ERVs and is associated with gene repression [19,20], whereas intragenic methylation prevents spurious transcription initiation [21]. Interestingly, human ERV-K (HERV-K) genomic sequences have undergone selection to mutate CpG dinucleotides, presumably to escape repression through DNA methylation [22]. *De novo* methylation is carried out by DNMT3A, DNMT3B, the newly discovered rodent-specific DNMT3C, and the cofactor DNMT3L [23[•],24–26] and maintained by DNMT1 through cell divisions [27,28]. Defective DNMT1 leads to DNA hypomethylation and ERV overexpression and plays a causative role in the onset of cancer and autoimmune disease [29,30]. Interestingly, the reactivation of ERVs observed following treatment of cells with DNA-demethylating agents, such as 5-azacytidine (5-AZA) [31] has been proposed to be responsible for driving anti-tumor immunity in cancer patients treated with these drugs through innate sensing of ERV nucleic acids [13^{••},14^{••}]. 5-AZA drugs are thought to target mainly DNMT3A and DNMT3B, which can be increased in expression upon differentiation [32], in contrast to DNMT3L, which is not expressed in differentiated cells [33].

Histone methyltransferases (HMTs) including SETDB1 (also known as ESET), SUV39h and G9a (also known as EHMT2) mediate retrotransposon repression through histone methylation [9,10,34]. The most relevant to this review is SETDB1 because it is required for retrotransposon repression in postembryonic tissues [35^{••},36]. Indeed, the finding that SETDB1 represses ERVs in committed mouse B-lineage cells, has led to a new paradigm that SETDB1 and potentially other histone modifiers remain important in differentiated cells [35^{••}]. Interestingly, loss of silent chromatin at SETDB1-regulated ERVs is not sufficient for their activation and the precise panel of ERVs reactivated in specific cell types depends on the transcription factors available [35^{••}]. HP1 too, which interacts with H3K9me3 [37,38] and participates in heterochromatin spreading [39] has been implicated in silencing of ERVs including in differentiated cells [40,41].

The KAP1 and KRAB-ZNF repertoire

The KAP1 and KRAB-ZNF (KRAB-zinc finger protein) system silences retrotransposons in early embryos and embryonic stem cells (ESCs) [12,42,43]. It has likely evolved in response to retrotransposon invasions [44] because KRAB-ZNF transcription factors are largely specific for transposon sequences [45–47] but now this pathway also participates in gene regulation [45,48[•],49]. KAP1 is recruited to transposons through the interaction of its RING, B-box and coiled-coil (RBCC) domain with the KRAB domain of KRAB-ZNFs. Transcriptional repression is mediated through co-factors such as SETDB1 [50–52], which prevents binding of transcriptional activators [53,54]. There are several hundred KRAB-ZNFs [55], which can be viewed as a panel of effector proteins specific for foreign DNA in the same way that an antibody repertoire is specific for foreign antigens. While many KRAB-ZNFs have recently been matched to their target sequences through chromatin immunoprecipitation experiments [45–47], only a few have been functionally characterized. These include human ZNF91 and ZNF93, which recognize specific SVA and LINE1 subfamily sequences, respectively [56[•]]. The KAP1 and KRAB-ZNF pathway is functional in human and mouse ESCs and neural progenitor cells [42,57–59] but little is known about its role in differentiated cells. However, one study showed that KAP1 binds to certain ERV-K elements in human primary CD4⁺ T cells [59]. It was also reported that KAP1 and KRAB-ZFPs bind to several ERVs in mouse liver. Interestingly, while KAP1 knockout in the liver had little impact on the expression of these ERVs, several co-regulated cellular genes were affected [48[•]]. This suggests that redundant mechanisms may converge to silence ERVs in differentiated cells [48[•]]. Surprisingly, HSP90 has recently been implicated in the formation of a KAP1 repressor complex at ERVs [60[•]]. Like KAP1, HSP90 is necessary to maintain silent chromatin at ERVs and prevent aberrant transcription of genes close to the ERVs that it regulates [49,60[•]]. Most interestingly, this is true not only in ESCs but also in differentiated macrophages [60[•]]. The nucleosomal and remodeling deacetylase (NuRD) complex, which interacts with KAP1 has also been implicated in retrotransposon repression in differentiated cells [61].

Of note, not all ERV-derived sequences are subject to epigenetic silencing as some have been co-opted because their non-coding DNA regulatory elements, nucleic acids or even gene products benefit their hosts [5,6,62^{••},63^{••}]. Whether co-opted ERVs are subject to spatial or temporal repression by KAP1 remains an open question. However, certain KRAB-ZNFs have been found to bind to co-opted ERVs and to recruit transcriptional activators and pioneer factors, suggesting that these ZNFs function to switch on rather than switch off certain gene networks [45,47].

Implications for immune regulation

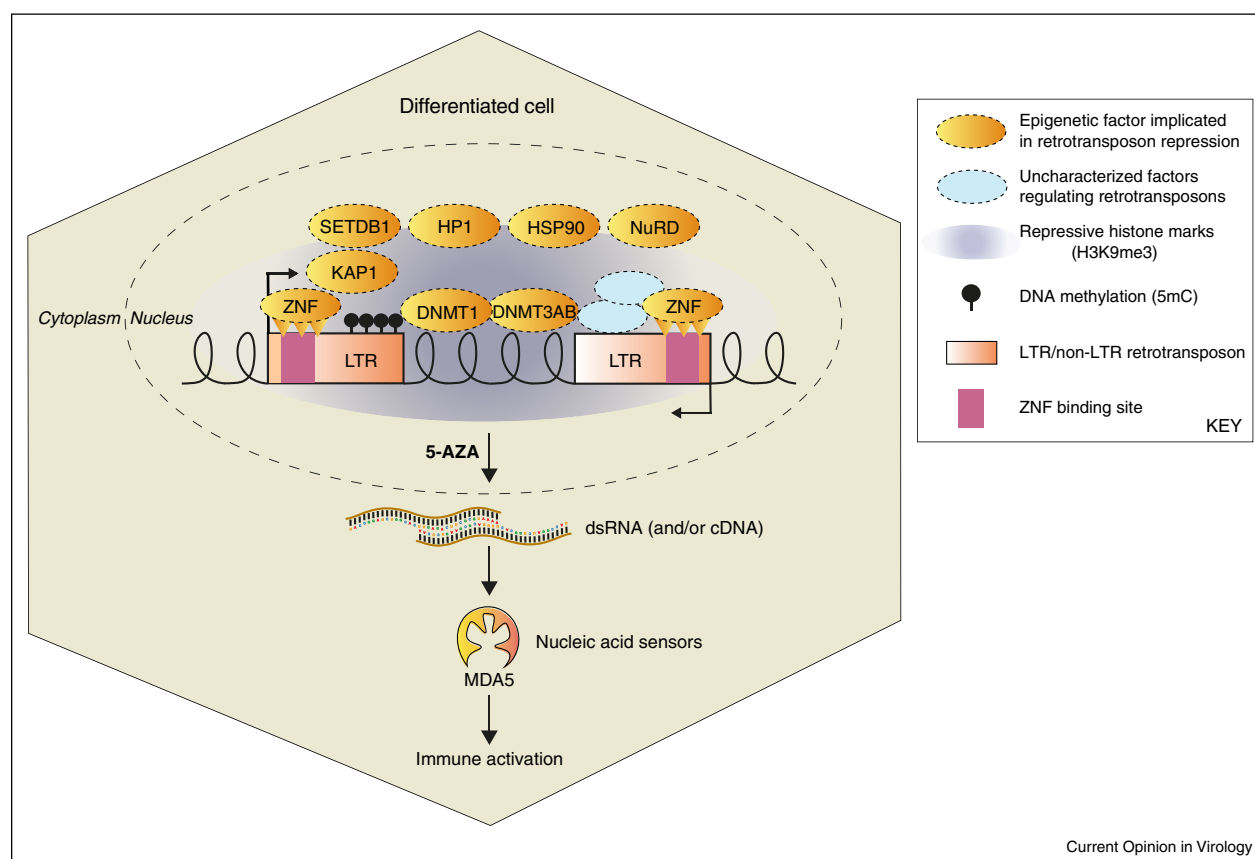
ERVs have been implicated in multiple cancers and autoimmune diseases, including ovarian and breast cancer, systemic lupus erythematosus and multiple sclerosis [64,65]. There is some convincing evidence that ERVs can play a causative role in cancer when their LTRs escape epigenetic repression [66] and interestingly, this involves mainly primate-specific ERVs [67]. For example, cryptic enhancers and promoters that reside within ERVs can drive expression of oncogenes [66,68*].

ERVs regulate the immune system in several ways. For example, double stranded RNA (dsRNA) produced from retrotransposons, following treatment of cancer cells with DNA methyltransferase inhibitors activates interferon through MDA5, MAVS and IRF7 [13**,14**]. Cytoplasmic DNA resulting from reverse transcription also serves as an additional PAMP because it is detected by the cytosolic DNA sensor cGAS to activate type I

interferon through STING [63**,69–72]. Likewise, Toll-like receptors contribute to ERV nucleic acid sensing [73]. Mutations within genes involved in nucleic acid metabolism including TREX1 are associated with autoimmune diseases [74], although such factors block classes of retrotransposons that are constitutively transcribed rather than those embedded within silent chromatin. In addition to innate immunity, ERVs can stimulate adaptive immunity too because their encoded gene products, which are necessary for their mobilization are subject to standard antigen processing and presentation pathways. For example, tumor-associated antigens (TAAs) can be derived from HERV-K envelope protein [75]. Such neo-antigens can evoke adaptive T cell and antibody responses [76], both of which have been demonstrated to regulate ERVs [15,75].

Overall, ERV regulatory sequences including solo LTRs have been described to contribute to cancer by driving

Figure 1



Epigenetic factors repress retrotransposons in differentiated tissues, preventing transcription and sensing of nucleic acids derived from endogenous viruses, which we refer to as 'non-self'. In differentiated cells, epigenetic silencing of retrotransposons is maintained by incompletely characterized mechanisms. Factors implicated in maintaining histone and DNA methylation at retrotransposons are shown. See text for details. Uncharacterized factors are depicted in blue to illustrate that this is an emerging area of research. The addition of 5-AZA-based drugs results in the generation of dsRNA from retrotransposons and potentially cDNA as well through reverse transcription, both of which lead to innate immune activation.

oncogenes, whereas longer ERVs if resurrected (e.g. through 5-AZA treatment) may promote anti-tumor immunity through their nucleic acids and proteins. Exactly how sensing of ERV nucleic acids leads to anti-tumor immunity is not fully understood but remarkably, it has been shown that if cancer initiating cells are pretreated with 5-AZA drugs before their injection into mice, they form less tumors and this phenotype is dependent on MAVS [13^{••},14^{••},68[•]]. Likewise, if B16 melanoma cells are pretreated with 5-AZA before their injection into mice that receive anti-CTLA-4, they can stimulate complete tumor clearance [13^{••}]. Furthermore, interferon-responsive genes are upregulated in cancer patients treated with 5-AZA [77[•]]. Interferon signaling is important presumably to promote apoptosis of cancer cells and to help to recruit cytotoxic T cells recognizing neo-antigens and other immune effectors to clear the tumor. It has also been shown that cytosolic RNA and DNA sensing of ERVs is necessary to induce T-independent B cell responses in mice [63^{••}]. This latter work illustrates that ERVs have coevolved with their hosts to play a natural role in modulating the immune system. Overall, ERVs lie at the intersection of innate and adaptive immunity, due to their intrinsic immunogenicity.

Concluding remarks

While it was previously thought that histone marks are primarily required to silence retrotransposons only early in development [10,50,78,79], where DNA methylation is reprogrammed [80,81], recent evidence has led to a new paradigm whereby diverse epigenetic modifiers exert continuous roles in adult tissues. Here we discuss evidence that SETDB1, DNMTs, HP1, HSP90, the NuRD complex and potentially KAP1 and KRAB-ZNFs are some of these factors (Figure 1). Importantly, most of these regulators have only been studied in mouse models so far. A future understanding of the pathways operating in adult human tissues is essential for the development of innovative drugs. Targeted epigenetic modulation might prove a potent tool in the future to reactivate certain retrotransposons so that their nucleic acids and proteins could serve as natural PAMPs to signal danger to their host. Such drugs may be valuable to stimulate immunosurveillance in cancer patients in which immune activation pathways may be subdued or could be used in conjunction with standard vaccines in place of an adjuvant. Caution should be applied, however, to prevent unwanted effects of reactivated ERVs on the genome or transcriptome.

Conflict of interest

The authors declare that they have no conflict of interests.

Acknowledgements

We thank Steen K Ooi, Greg Towers, Pierre V Maillard, Luisa Robbez-Masson and Connor Husovsky for helpful comments. This work was supported through a Sir Henry Dale Fellowship jointly funded by the

Wellcome Trust and Royal Society (Grant number 101200/Z/13/Z) awarded to HMR and a UCL Grand Challenges Studentship (number 518099) and UCL Overseas Research Scholarship both awarded to CHCT.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Mc CB: **The origin and behavior of mutable loci in maize.** *Proc Natl Acad Sci U S A* 1950, **36**:344-355.
2. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD: **Repetitive elements may comprise over two-thirds of the human genome.** *PLoS Genet* 2011, **7**:e1002384.
3. Henzy JE, Gifford RJ, Kenaley CP, Johnson WE: **An intact retroviral gene conserved in spiny-rayed fishes for over 100 My.** *Mol Biol Evol* 2017, **34**:634-639.
4. Aiewsakun P, Katzourakis A: **Marine origin of retroviruses in the early Palaeozoic Era.** *Nat Commun* 2017, **8**:13954.
5. Dupressoir A, Vernochet C, Bawa O, Harper F, Pierron G, Opolon P, Heidmann T: **Syncytin-A knockout mice demonstrate the critical role in placentalization of a fusogenic, endogenous retrovirus-derived, envelope gene.** *Proc Natl Acad Sci U S A* 2009, **106**:12127-12132.
6. Wang J, Xie G, Singh M, Ghanbarian AT, Rasko T, Szvetnik A, Cai H, Besser D, Prigione A, Fuchs NV *et al.*: **Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells.** *Nature* 2014, **516**:405-409.
7. Shukla R, Upton KR, Munoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, Brennan PM, Baillie JK, Collino A, Ghisletti S *et al.*: **Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma.** *Cell* 2013, **153**:101-111.
8. Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J: **A unique chromatin signature uncovers early developmental enhancers in humans.** *Nature* 2011, **470**:279-283.
9. Bulut-Karslioglu A, De La Rosa-Velazquez IA, Ramirez F, Barenboim M, Onishi-Seebacher M, Arand J, Galan C, Winter GE, Engist B, Gerle B *et al.*: **Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells.** *Mol Cell* 2014, **55**:277-290.
10. Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y: **Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET.** *Nature* 2010, **464**:927-931.
11. Walter M, Teissandier A, Perez-Palacios R, Bourc'his D: **An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells.** *Elife* 2016:5.
12. Wolf D, Goff SP: **Embryonic stem cells use ZFP809 to silence retroviral DNAs.** *Nature* 2009, **458**:1201-1204.
13. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, Hein A, Rote NS, Cope LM, Snyder A *et al.*: **Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses.** *Cell* 2015, **162**:974-986.
- This and Ref. [14^{••}] are the first reports implicating ERVs in the mechanism of anti-tumour immunity induced by 5-AZA treatment in cancer patients.
14. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, Han H, Liang G, Jones PA, Pugh TJ *et al.*: **DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts.** *Cell* 2015, **162**:961-973.
- See annotation to Ref. [13^{••}].
15. Young GR, Eksmond U, Salcedo R, Alexopoulou L, Stoye JP, Kassiotis G: **Resurrection of endogenous retroviruses in antibody-deficient mice.** *Nature* 2012, **491**:774-778.
16. Goodier JL: **Restricting retrotransposons: a review.** *Mobile DNA* 2016, **7**:16.

17. Wolf G, Yang P, Fuchtbauer AC, Fuchtbauer EM, Silva AM, Park C, Wu W, Nielsen AL, Pedersen FS, Macfarlan TS: **The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses.** *Genes Dev* 2015, **29**:538-554.
- The first report of a KRAB-ZFP knockout mouse revealing that absence of Zfp809 in early embryos leads to the upregulation of its target ERVs in differentiated tissues.
18. Campos-Sanchez R, Cremona MA, Pini A, Chiaromonte F, Makova KD: **Integration and fixation preferences of human and mouse endogenous retroviruses uncovered with functional data analysis.** *PLoS Comput Biol* 2016, **12**:e1004956.
19. Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A: **Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice.** *Oncogene* 2008, **27**:404-408.
20. Walsh CP, Chaillet JR, Bestor TH: **Transcription of IAP endogenous retroviruses is constrained by cytosine methylation.** *Nat Genet* 1998, **20**:116-117.
21. Neri F, Rapelli S, Krepelova A, Incarnato D, Parlato C, Basile G, Maldotti M, Anselmi F, Oliviero S: **Intragenic DNA methylation prevents spurious transcription initiation.** *Nature* 2017, **543**:72-77.
22. Zsiros J, Jebbink MF, Lukashov VV, Voute PA, Berkhout B: **Biased nucleotide composition of the genome of HERV-K related endogenous retroviruses and its evolutionary implications.** *J Mol Evol* 1999, **48**:102-111.
23. Barau J, Teissandier A, Zamudio N, Roy S, Nalesso V, Herault Y, Guillou F, Bourc'his D: **The DNA methyltransferase DNMT3C protects male germ cells from transposon activity.** *Science* 2016, **354**:909-912.
- Identification of a novel DNMT gene that represses transposons.
24. Chedin F, Lieber MR, Hsieh CL: **The DNA methyltransferase-like protein DNMT3L stimulates de novo methylation by Dnmt3a.** *Proc Natl Acad Sci U S A* 2002, **99**:16916-16921.
25. Okano M, Bell DW, Haber DA, Li E: **DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development.** *Cell* 1999, **99**:247-257.
26. Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD *et al.*: **DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA.** *Nature* 2007, **448**:714-717.
27. Hutnick LK, Huang X, Loo TC, Ma Z, Fan G: **Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism.** *J Biol Chem* 2010, **285**:21082-21091.
28. Liao J, Karnik R, Gu H, Ziller MJ, Clement K, Tsankov AM, Akopian V, Gifford CA, Donaghey J, Galonska C *et al.*: **Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells.** *Nat Genet* 2015, **47**:469-478.
29. Fali T, Le Dantec C, Thabet Y, Jousse S, Hanrotel C, Youinou P, Brooks WH, Perl A, Renaudineau Y: **DNA methylation modulates HRES1/p28 expression in B cells from patients with Lupus.** *Autoimmunity* 2014, **47**:265-271.
30. Gaudet F, Rideout WM 3rd, Meissner A, Dausman J, Leonhardt H, Jaenisch R: **Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing.** *Mol Cell Biol* 2004, **24**:1640-1648.
31. Groudine M, Eisenman R, Weintraub H: **Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation.** *Nature* 1981, **292**:311-317.
32. Oka M, Meacham AM, Hamazaki T, Rodic N, Chang LJ, Terada N: **De novo DNA methyltransferases Dnmt3a and Dnmt3b primarily mediate the cytotoxic effect of 5-aza-2'-deoxycytidine.** *Oncogene* 2005, **24**:3091-3099.
33. Kao TH, Liao HF, Wolf D, Tai KY, Chuang CY, Lee HS, Kuo HC, Hata K, Zhang X, Cheng X *et al.*: **Ectopic DNMT3L triggers assembly of a repressive complex for retroviral silencing in somatic cells.** *J Virol* 2014, **88**:10680-10695.
34. Leung DC, Dong KB, Maksakova IA, Goyal P, Appanah R, Lee S, Tachibana M, Shinkai Y, Lehnertz B, Mager DL *et al.*: **Lysine methyltransferase G9a is required for de novo DNA methylation and the establishment, but not the maintenance, of proviral silencing.** *Proc Natl Acad Sci U S A* 2011, **108**:5718-5723.
35. Collins PL, Kyle KE, Egawa T, Shinkai Y, Oltz EM: **The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes.** *Proc Natl Acad Sci U S A* 2015, **112**:8367-8372.
- This paper introduces a new paradigm that histone methyltransferases and potentially other epigenetic modifiers are required for retrotransposon repression in differentiated cells.
36. Pasquarella A, Ebert A, Pereira de Almeida G, Hinterberger M, Kazerani M, Nuber A, Ellwart J, Klein L, Busslinger M, Schotta G: **Retrotransposon derepression leads to activation of the unfolded protein response and apoptosis in pro-B cells.** *Development* 2016, **143**:1788-1799.
37. Hiragami-Hamada K, Soeroes S, Nikolov M, Wilkins B, Kreuz S, Chen C, De La Rosa-Velazquez IA, Zenn HM, Kost N, Pohl W *et al.*: **Dynamic and flexible H3K9me3 bridging via HP1beta dimerization establishes a plastic state of condensed chromatin.** *Nat Commun* 2016, **7**:11310.
38. Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T: **Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins.** *Nature* 2001, **410**:116-120.
39. Canzio D, Liao M, Naber N, Pate E, Larson A, Wu S, Marina DB, Garcia JF, Madhani HD, Cooke R *et al.*: **A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly.** *Nature* 2013, **496**:377-381.
40. Sharma P, Azebi S, England P, Christensen T, Moller-Larsen A, Petersen T, Batsche E, Muchardt C: **Citrullination of histone H3 interferes with HP1-mediated transcriptional repression.** *PLoS Genet* 2012, **8**:e1002934.
41. Wolf D, Cammas F, Losson R, Goff SP: **Primer binding site-dependent restriction of murine leukemia virus requires HP1 binding by TRIM28.** *J Virol* 2008, **82**:4675-4679.
42. Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J *et al.*: **KAP1 controls endogenous retroviruses in embryonic stem cells.** *Nature* 2010, **463**:237-240.
43. Wolf D, Goff SP: **TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells.** *Cell* 2007, **131**:46-57.
44. Thomas JH, Schneider S: **Coevolution of retroelements and tandem zinc finger genes.** *Genome Res* 2011, **21**:1800-1812.
45. Imbeault M, Helleboid PY, Trono D: **KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks.** *Nature* 2017, **543**:550-554.
46. Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A, Albu M, Weirauch MT, Radovani E, Kim PM *et al.*: **C2H2 zinc finger proteins greatly expand the human regulatory lexicon.** *Nat Biotechnol* 2015, **33**:555-562.
47. Schmitges FW, Radovani E, Najafabadi HS, Barazandeh M, Campitelli LF, Yin Y, Jolma A, Zhong G, Guo H, Kanagalingam T *et al.*: **Multiparameter functional diversity of human C2H2 zinc finger proteins.** *Genome Res* 2016, **26**:1742-1752.
48. Ecco G, Cassano M, Kauzlaric A, Duc J, Coluccio A, Offner S, Imbeault M, Rowe HM, Turelli P, Trono D: **Transposable elements and their KRAB-ZFP controllers regulate gene expression in adult tissues.** *Dev Cell* 2016, **36**:611-623.
- First report of KAP1 and KRAB-ZFP repression of ERVs in differentiated mouse tissues.
49. Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, Tarabay Y, Viville S, Jakobsson J, Pfaff SL, Trono D: **TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells.** *Genome Res* 2013, **23**:452-461.
50. Elsasser SJ, Noh KM, Diaz N, Allis CD, Banaszyński LA: **Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells.** *Nature* 2015, **522**:240-244.
51. Leung D, Du T, Wagner U, Xie W, Lee AY, Goyal P, Li Y, Szulwach KE, Jin P, Lorincz MC *et al.*: **Regulation of DNA**

- methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1.** *Proc Natl Acad Sci U S A* 2014, **111**:6690-6695.
52. Yang BX, El Farran CA, Guo HC, Yu T, Fang HT, Wang HF, Schlesinger S, Seah YF, Goh GY, Neo SP *et al.*: **Systematic identification of factors for provirus silencing in embryonic stem cells.** *Cell* 2015, **163**:230-245.
 53. de la Rica L, Deniz O, Cheng KC, Todd CD, Cruz C, Houseley J, Branco MR: **TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells.** *Genome Biol* 2016, **17**:234.
 54. Ohnuki M, Tanabe K, Sutou K, Teramoto I, Sawamura Y, Narita M, Nakamura M, Tokunaga Y, Nakamura M, Watanabe A *et al.*: **Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential.** *Proc Natl Acad Sci U S A* 2014, **111**:12426-12431.
 55. Corsinotti A, Kapopoulou A, Gubelmann C, Imbeault M, Santoni de Sio FR, Rowe HM, Mouscay Y, Deplancke B, Trono D: **Global and stage specific patterns of Kruppel-associated-box zinc finger protein gene expression in murine early embryonic cells.** *PLoS One* 2014, **8**:e56721.
 56. Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D: **An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons.** *Nature* 2014, **516**:242-245.
- First functional and evolutionary characterization of human KRAB-ZNFs recognizing retrotransposons.
57. Brattas PL, Jonsson ME, Fasching L, Nelander Wahlestedt J, Shahsavani M, Falk R, Falk A, Jern P, Parmar M, Jakobsson J: **TRIM28 controls a gene regulatory network based on endogenous Retroviruses in human neural progenitor cells.** *Cell Rep* 2017, **18**:1-11.
 58. Fasching L, Kapopoulou A, Sachdeva R, Petri R, Jonsson ME, Manne C, Turelli P, Jern P, Cammas F, Trono D *et al.*: **TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells.** *Cell Rep* 2015, **10**:20-28.
 59. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, Tieng V, Quenneville S, Trono D: **Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements.** *Genome Res* 2014, **24**:1260-1270.
 60. Hummel B, Hansen EC, Yoveva A, Aprile-Garcia F, Hussong R, Sawarkar R: **The evolutionary capacitor HSP90 buffers the regulatory effects of mammalian endogenous retroviruses.** *Nat Struct Mol Biol* 2017, **24**:234-242.
- HSP90 is required to maintain ERVs repressed in differentiated mouse macrophages.
61. Montoya-Durango DE, Ramos KA, Bojang P, Ruiz L, Ramos IN, Ramos KS: **LINE-1 silencing by retinoblastoma proteins is effected through the nucleosomal and remodeling deacetylase multiprotein complex.** *BMC Cancer* 2016, **16**:38.
 62. Chuong EB, Elde NC, Feschotte C: **Regulatory evolution of innate immunity through co-option of endogenous retroviruses.** *Science* 2016, **351**:1083-1087.
- The first report that ERV regulatory DNA has been co-opted in the normal functioning of the immune system.
63. Zeng M, Hu Z, Shi X, Li X, Zhan X, Li XD, Wang J, Choi JH, Wang KW, Purrington T *et al.*: **MAVS, cGAS, and endogenous retroviruses in T-independent B cell responses.** *Science* 2014, **346**:1486-1492.
- This report suggests that ERV nucleic acids have been co-opted into the normal functioning of the immune system.
64. Menendez L, Benigno BB, McDonald JF: **L1 and HERV-W retrotransposons are hypomethylated in human ovarian carcinomas.** *Mol Cancer* 2004, **3**:12.
 65. Wang-Johanning F, Li M, Esteva FJ, Hess KR, Yin B, Rycaj K, Plummer JB, Garza JG, Ambs S, Johanning GL: **Human endogenous retrovirus type K antibodies and mRNA as serum biomarkers of early-stage breast cancer.** *Int J Cancer* 2014, **134**:587-595.
 66. Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenze D, Kochert K, Bouhrel MA, Richter J, Soler E *et al.*: **Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma.** *Nat Med* 2010, **16**:571-579 571p following 579.
 67. Szpakowski S, Sun X, Lage JM, Dyer A, Rubinstein J, Kowalski D, Sasaki C, Costa J, Lizardi PM: **Loss of epigenetic silencing in tumors preferentially affects primate-specific retroelements.** *Gene* 2009, **448**:151-167.
 68. Babaian A, Romanish MT, Gagnier L, Kuo LY, Karimi MM, Steidl C, Mager DL: **Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma.** *Oncogene* 2016, **35**:2542-2546.
- Recent demonstration that cryptic LTRs can drive oncogenes in human cancers.
69. Gao D, Wu J, Wu YT, Du F, Aroh C, Yan N, Sun L, Chen ZJ: **Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses.** *Science* 2013, **341**:903-906.
 70. Hartlova A, Ertmann SF, Raffi FA, Schmalz AM, Resch U, Anugula S, Lienenklaus S, Nilsson LM, Kroger A, Nilsson JA *et al.*: **DNA damage primes the type I interferon system via the cytosolic DNA sensor STING to promote anti-microbial innate immunity.** *Immunity* 2015, **42**:332-343.
 71. Rasaiyaah J, Tan CP, Fletcher AJ, Price AJ, Blondeau C, Hilditch L, Jacques DA, Selwood DL, James LC, Noursadeghi M *et al.*: **HIV-1 evades innate immune recognition through specific cofactor recruitment.** *Nature* 2013, **503**:402-405.
 72. Sun L, Wu J, Du F, Chen X, Chen ZJ: **Cyclic GMP-AMP synthase is a cytosolic DNA sensor that activates the type I interferon pathway.** *Science* 2013, **339**:786-791.
 73. Yu P, Lubben W, Slomka H, Gebler J, Konert M, Cai C, Neubrandt L, Prazeres da Costa O, Paul S, Dehnert S *et al.*: **Nucleic acid-sensing Toll-like receptors are essential for the control of endogenous retrovirus viremia and ERV-induced tumors.** *Immunity* 2012, **37**:867-879.
 74. Gray EE, Winship D, Snyder JM, Child SJ, Geballe AP, Stetson DB: **The AIM2-like receptors are dispensable for the interferon response to intracellular DNA.** *Immunity* 2016, **45**:255-266.
 75. Krishnamurthy J, Rabinovich BA, Mi T, Switzer KC, Olivares S, Maiti SN, Plummer JB, Singh H, Kumaresan PR, Huls HM *et al.*: **Genetic engineering of T cells to target HERV-K, an ancient Retrovirus on melanoma.** *Clin Cancer Res: Off J Am Assoc Cancer Res* 2015, **21**:3241-3251.
 76. Takahashi Y, Harashima N, Kajigaya S, Yokoyama H, Cherkasova E, McCoy JP, Hanada K, Mena O, Kurlander R, Tawab A *et al.*: **Regression of human kidney cancer following allogeneic stem cell transplantation is associated with recognition of an HERV-E antigen by T cells.** *J Clin Invest* 2008, **118**:1099-1109.
 77. Li H, Chiappinelli KB, Guzzetta AA, Easwaran H, Yen RW, Vatapalli R, Topper MJ, Luo J, Connolly RM, Azad NS *et al.*: **Immune regulation by low doses of the DNA methyltransferase inhibitor 5-azacitidine in common human epithelial cancers.** *Oncotarget* 2014, **5**:587-598.
- Demonstration that interferon-responsive genes are upregulated in patients receiving 5-AZA treatment.
78. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP *et al.*: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.** *Nature* 2007, **448**:553-560.
 79. Rowe HM, Friedli M, Offner S, Verp S, Mesnard D, Marquis J, Aktas T, Trono D: **De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET.** *Development* 2013, **140**:519-529.
 80. Iqbal K, Jin SG, Pfeifer GP, Szabo PE: **Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine.** *Proc Natl Acad Sci U S A* 2011, **108**:3642-3647.
 81. Ficiz G, Hore TA, Santos F, Lee HJ, Dean W, Arand J, Krueger F, Oxley D, Paul YL, Walter J *et al.*: **FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency.** *Cell Stem Cell* 2013, **13**:351-359.

Appendix B

“Under revision for *EMBO Reports*”

KAP1 regulates ERVs in differentiated human cells and contributes to the control of viral mimicry

Christopher H.C. Tie¹, Luisa Robbez-Masson¹, Lucia Conde², Maria Teresa Rodriguez-Plata¹, Greg J. Towers¹, Robert Gifford³, Javier Herrero² and Helen M. Rowe¹

¹Division of Infection and Immunity, Cruciform Building, University College London, WC1E 6BT, UK

²Bill Lyons Informatics Centre, UCL Cancer Institute, Paul O’Gorman Building, 72 Huntley Street, London WC1E 6BT, UK

³MRC-University of Glasgow Centre for Virus Research, G61 1QH, UK

Corresponding author: h.rowe@ucl.ac.uk

Running title: **KAP1 role and relevance in human cells**

Abstract

Endogenous retroviruses (ERVs) have accumulated in vertebrate genomes and contribute to the complexity of gene regulation. KAP1 represses ERVs during development by its recruitment to their repetitive sequences through Krüppel-associated box domain-zinc finger proteins (KZNFs), but little is known about the regulation of ERVs in differentiated cells. We observed that KAP1 repression of HERVK14C was conserved in differentiated human cells and performed KAP1 knockout and mRNA-sequencing to obtain an overview of KAP1 function. Our results show that KAP1 represses ERVs and ZNFs, both of which overlap with KAP1 binding sites and silent chromatin marks in multiple cell types. Furthermore, this pathway is functionally conserved in primary human peripheral blood mononuclear cells (PBMCs). Finally, we reveal that cytosine methylation that acts on KAP1-regulated loci is necessary to prevent innate immune reactivity of ERVs and other retrotransposons, which can mimic viruses by producing immunostimulatory nucleic acids. These data indicate that the KAP1-KZNF pathway has evolved to play an important functional role in genome stability and the control of viral mimicry in differentiated human cells.

Keywords: epigenetic control / human differentiated cells / endogenous retroviruses (ERVs) / nucleic acid sensing / KAP1 (KRAB-associated protein 1)

Introduction

Retrotransposons are genetic elements capable of amplifying their copy number in the genome by replicating through an RNA intermediate. They now occupy at least half of the human genome and individual loci reflect the genetic conflict experienced by the host throughout evolution[1, 2]. Retrotransposons are divided into two groups, those with long terminal repeats (LTR) and those without (non-LTR). LTR retrotransposons are known as endogenous retroviruses (ERVs) because their genetic make-up is like that of exogenous retroviruses from which they are derived. ERVs constitute around 9% of the human genome[3, 4] and are a class of retrotransposons with important roles in human health and disease. For example, LTRs of the primate-specific retrovirus, MER41, have been co-opted to function as natural poised enhancers for a network of interferon-induced genes[5]. The co-option of ERVs into normal processes such as development and immune function is the result of co-evolution of ERVs and their regulatory DNA sequences, which are scattered across the genome, with their hosts over millions of years[6].

ERVs are subject to epigenetic repression in adult tissues involving mechanisms including their transcriptional repression through cytosine methylation [7]. This is critical to prevent expression of their nucleic acids and proteins, which could potentially trigger an autoimmune response. This has recently been illustrated by two studies in which treatment of cancer cells with 5-AZA-based DNA demethylating agents led to the reactivation of ERVs[8, 9]. This in turn activated interferon-

stimulated genes (ISGs) through MDA-5, MAVS and IRF7 following sensing of double stranded RNA transcripts, including those derived from ERVs. The authors propose that this mechanism contributes to anti-tumour immunity in patients treated with these drugs and remarkably, cancer initiating cells pretreated with 5-AZA form less tumours in mice in a MAVS-dependent manner[9]. Like-wise, reactivation of ERVs in differentiated cells has also been linked to autoimmune diseases such as multiple sclerosis[10]. Despite the presence and importance of ERV transcriptional regulation in adult tissues, however, very little is known about the epigenetic control pathways in operation. In contrast, the mechanisms in place to restrict ERVs early in development are well-described. In early embryos, in which DNA methylation is erased, epigenetic complexes act to establish histone and cytosine methylation patterns at ERVs[11]. This serves to protect transcriptome integrity and prevent vertical transmission of retrotransposition events. One such pathway involves KAP1 and krüppel-associated box domain-zinc finger proteins (KZNFs)[12-16]. KZNFs (known as KZFPs in mouse) interact with their corresponding DNA targets, which are mainly located within repetitive DNA [17-19], and recruit KAP1 to these loci[20-22]. Gene silencing ensues when KAP1 binds epigenetic modifiers including HP1, SETDB1 and DNMTs [23-25].

KAP1 has been shown to bind to certain ERVs in human CD4+ T cells[13] and plays a functional role in ERV regulation in human neural progenitor cells, although it is not clear if this is the case in differentiated neurons[26]. KAP1 has been depleted in several differentiated murine cell types with little effect on ERVs[27, 28]. Interestingly, however, its ablation in mouse liver results in the overexpression of

several cellular genes close to KAP1-bound ERVs, although the ERVs themselves remain repressed[27]. In contrast, SETDB1, which is a binding partner of KAP1 has been shown to be necessary for the silencing of ERVs in mouse B lymphocytes[29].

Therefore, the potential role and relevance of KAP1 in differentiated cells, particularly in humans, is a fundamental and open question. We set out to address this question by employing CRISPR/Cas9-mediated KAP1 knockout in human differentiated cells followed by mRNA-sequencing and correlated our findings with KAP1 binding data. We report here that KAP1 binds to and represses ERVs and ZNFs in differentiated human cells. We demonstrate using reporters that KAP1 can induce *de novo* ERV repression even in primary human peripheral blood mononuclear cells (PBMCs). Finally, we show evidence that KAP1 is one epigenetic regulator that controls retrotransposon nucleic acid sensing and activation of interferon-stimulated genes. These results reveal the widespread importance of the KAP1 pathway in genome stability and the regulation of viral mimicry in differentiated human cells. This work also illustrates that cell lines could be used to dissect the functions of KAP1-recruiting KZNFs in human gene regulation through their targeted gene knockout.

Results

The human retrovirus HERVK14C is repressed by KAP1 in undifferentiated and differentiated cells

KAP1 is known to regulate the human ERV (HERV) lineage HERVK14C in undifferentiated human embryonic stem cells (ESCs)[13]. We therefore selected HERVK-14C as a tool to explore KAP1 function in other cell types. HERVK14C is restricted to Great Apes and Old World Monkeys (Figure 1a) and is a low copy number ERV making it relatively easy to study. We first identified all loci that included internal regions (Figure 1b) because KAP1 has been documented to frequently target internal ERV sites for repression including sequences overlapping the primer binding site (PBS), GAG and ENV[15, 27, 30]. We designed two primer sets to detect specific HERVK14C loci (Figure 1c) and employed qRT-PCR to confirm that KAP1 represses these ERVs in undifferentiated Oct4-expressing embryonic NTERA-2 cells (Figure 1d and Figure S1a). We then focused on two differentiated cell lines, HeLa and 293T cells and generated KAP1 knockout clones using CRISPR/Cas9 genome editing. We validated KAP1 knockout and complementation functionally using described KAP1-KZNF reporters[14, 16] (Figure S1bc) and we discovered that like in undifferentiated cells, KAP1 represses HERVK14C in differentiated cell lines (Figure 1ef). The phenotype was consistent between clones, although the magnitude of the effect was variable and we also observed differences between primer sets, which recognize overlapping as well as distinct loci (Figure 1cef). Two other KAP1-regulated retrotransposons, L1PA4 (a LINE1 subfamily) and SVA D (a SVA subfamily), in contrast, were only modestly affected by KAP1 depletion in both undifferentiated

and differentiated cells (Figure 1def). This may reflect redundant silencing mechanisms at these elements and in line with this, we found SVA elements to often be enriched for cytosine methylation (Figure S1d).

A common role for KAP1 in repressing ERVs and ZNFs

Studies exploring KAP1 function have been hindered by the fact that complete KAP1 knockout is usually lethal, while knockdown may not be sufficient to reveal a phenotype in all cell types [12, 13, 31]. Since here we could obtain KAP1 knockout HeLa cell clones (Figure 1), we employed mRNA-sequencing to gain a systematic overview of the role of KAP1 in these differentiated cells. Sequencing reads were mapped to the human genome and to RepBase to assess the global expression of genes and repetitive elements respectively, and we focused on significant changes (where differential effects were > 2 -fold with p -values < 0.05). Results reveal that KAP1-knockout induces a very specific and modest phenotype involving overexpression of ERVs and ZNFs (Figure 2). HERVs (HERV-T, HERV-S and HERVK14C) were the only class of repetitive elements overexpressed in knockout HeLa cells (Figure 2a and Spreadsheet 1) and as expected, these ERV families were also bound by KAP1 according to ENCODE data (Figure 2a and S2a). Intriguingly, these ERVs are present in diverse primates and other mammals (Figure 2b). Of note, several retrotransposons were downregulated in knockout cells (Figure S2b). The fact that few ERVs were reactivated may relate to the low number of active ERV transcriptional units in the human compared to the mouse genome[4]. Interrogation of the transcriptome showed that KAP1 knockout also affects several hundred cellular genes (Figure S3ab, Spreadsheet 2). When we focused on upregulated genes

(>2fold where $p_{adj} \leq 0.05$), we found that strikingly, ZNFs, of which there were 13 (9 of which were KZNFs) are the class of cellular genes most significantly overexpressed (Figure 2c). All of these KZNFs are also direct KAP1-binding targets (using ENCODE data)(Figure S3c), illustrating that KAP1 plays a functional role in regulating these sites, a notion that was previously inferred only from binding rather than functional data[32, 33].

Since KAP1 repression of ERVs and ZNFs is a phenotype that has also been observed in embryonic cells[12, 13], we hypothesized that KAP1 may bind ERVs and ZNFs at common sites in both undifferentiated and differentiated cells. To address this question, we determined if any KAP1 binding sites were common between human ESCs and differentiated cells (293T cells) using public ChIP-seq data ([13]and ENCODE data and identified 614 common peaks (Figure S3d and Spreadsheet 3). We found these loci to be highly enriched for ERVs compared to their abundance in the genome (Figure 2d). We determined the nearest gene to each of these sites and interestingly, gene ontology analysis showed that the most common gene cluster was ZNFs, (of which there were 61, including 40 KZNFs) (Figure 2e), mirroring our functional data for the upregulated genes (Figure 2c). Finally, examining the occurrence of LINE1 elements and ERVs within KAP1 binding sites in (i) human ESCs or in (ii) 293T cells or (iii) within common sites revealed a similar pattern of distribution, except that ERV1 elements were slightly enriched within the common sites (Figure 2f). This suggests that the landscape of KAP1 binding does not change dramatically between these cell types. The 614 common peaks are likely particularly strong KAP1-binding targets allowing their consistent identification. Overall, these

results illustrate that ERVs and ZNFs are repressed by KAP1 through direct binding in undifferentiated and differentiated cells.

KAP1 common binding sites are enriched for repressive histone marks in multiple cell types and overlap KZNF ERV targets

We hypothesized that KAP1 repression of key ERV and ZNF loci may be a common feature in multiple cell types. To address this question, we probed ENCODE data of epigenetic marks in undifferentiated, differentiated and primary cell types and asked if the common KAP1 binding sites we identified (Figure S3d) overlapped silent chromatin marks (Figure 3). We found indeed that these loci correlate with KAP1 and SETDB1 binding and repressive H3K9me3 in multiple cell types (Figure 3a). In comparison, we did not observe a correlation with the active chromatin marks, H3K4me1 and H3K27ac (Figure 3a) as expected. KAP1 is recruited to genomic sites through KZNFs[14, 16, 32], which we show here to be maintained at low expression levels through KAP1 binding (Figure 2). We reasoned that a subset of KZNFs must be widely expressed and at a sufficient level to recruit KAP1 to common sites. We therefore identified the top one hundred most highly expressed KZNFs in 293T cells using mRNA-sequencing data[34], verified them to be expressed at the protein level (the human protein atlas[35]) and recorded their targets where known[17, 19] (Figure 3b and Spreadsheet 4). Many of these KZNFs bind to ERVs (Figure 3b), reflecting KAP1 binding at common loci (Figure 3c) and interestingly, the ERV-binding KZNFs are widely conserved among primates or mammals and recognize ERVL or ERV1 sequences (Figure 3d). These data show that KAP1 and KZNFs repress ERVs and ZNFs in multiple cell types and suggest that core interactions may even be conserved

between species.

KAP1 restricts incoming ERVs through the primer binding site sequence

We next asked if differentiated cells were equipped with high enough levels of KAP1 and its associated epigenetic machinery to induce *de novo* repression of incoming ERV sequences. KAP1 is known to target several retroviruses including HERVK through the primer binding site (PBS) [13, 15, 36] as well as other sites[27, 30]. We selected the KAP1-bound and repressed HERVK14C integrant on chromosome 15 as a model to measure *de novo* KAP1 repression in a reporter assay. It contains a variant lysine PBS sequence (PBSChr15), which we cloned upstream of the HERVK14C LTR in a vector (termed PBSChr15-LTR-GFP) and tested alongside a vector containing the consensus PBS-lys1,2 sequence (PBS-LTR-GFP) or a control vector with no PBS (LTR-GFP) (Figure 4a). Using this system, we could demonstrate that the HERVK14C chromosome 15 PBS sequence and the consensus PBS sequence could induce KAP1-dependent repression in both undifferentiated and differentiated cells (Figure 4bc). In contrast, the control vector was not subject to significant KAP1 repression (Figure 4bc).

PBS-dependent *de novo* repression of ERVs is conserved in primary cells

A key question is whether the KAP1 pathway is conserved and functional in repressing ERVs in differentiated primary human cells. Existing data suggests that KAP1 binding is conserved at specific ERVs in primary CD4+ T cells[13]. Here, we were interested in the relevance of the KAP1 pathway in controlling the potential immune response elicited by uncontrolled ERV expression so we focused on PBMCs

as a model because they participate in innate and adaptive immunity. We found that KAP1 depletion in PBMCs leads to overexpression of HERVK14C and other KAP1-repressed retrotransposons (Figure 5a and S4a), consistent with the overlap of KAP1 binding and silent chromatin at ERVs in PBMCs (Figure 3). Parallel experiments in activated CD4+ T cells induced a more modest effect (Figure S4bc), suggesting that the KAP1 pathway may be partly redundant with other epigenetic pathways in primary cells. Indeed, we could detect cytosine methylation at the HERVK14C LTR in CD4+ T cells (Figure S4d). In HeLa cells, interestingly, cytosine methylation was reduced at HERVK14C by KAP1 knockout suggesting that KAP1 maintains this mark (Figure S4e). It is possible that sustained KAP1 knockout in primary cells would lead to more pronounced overexpression of ERVs and a decrease in cytosine methylation but we were unable to generate stable KAP1-depleted primary cells due to toxicity. To overcome this problem, we used our HERVK reporters (Figure 4), which are themselves unmethylated to transduce PBMCs and measure the establishment of repression. Results showed that the PBS-containing vector was repressed relative to the control vector (Figure 5b), while both were integrated at similar proviral copy numbers. This data demonstrates that *de novo* KAP1-mediated transcriptional repression is intact in PBMCs.

We hypothesized that a core set of KZNFs must be expressed in PBMCs as well as in other cell types to mediate KAP1 repression of key sequences. To explore this possibility, we used mRNA-sequencing data (see methods for accession numbers) and selected all KZNFs that are expressed within a given cell type and looked at the overlap of these KZNFs between cell types. Results show that seventy-seven KZNFs are widely expressed at the mRNA level (Figure 5c, Spreadsheet 5). Finally, we

verified that several of these KZNFs are also expressed at the protein level (Figure 5d) and we checked that the top one hundred KZNFs in 293T cells (Spreadsheet 4) have been described to be expressed at the protein level[35].

Cytosine methylation acts on KAP1-regulated retrotransposons and prevents viral mimicry and innate immune activation

ERVs are subject to epigenetic repression in cancer cells and disruption of cytosine methylation with 5-AZA-based drugs leads to the activation of ISGs through ERV-derived nucleic acid sensing [8, 9]. However, while the signaling pathways that 5-AZA affects are well-established, the identities of the epigenetic factors that repress 5-AZA-modulated ERVs are unknown. It's also unclear whether 5-AZA acts specifically on cancer cells or if it can reactivate ERVs in normal differentiated cells too. We reasoned that KAP1 may be one epigenetic player relevant to the 5-AZA affect because (i) it represses ERVs in differentiated cells and (ii) it recruits SETDB1, H3K9me3 and DNA methylation [13, 24, 25, 37, 38]. To explore this possibility, we performed experiments in cell lines and PBMCs where we depleted KAP1 or treated cells with 5-AZA and measured derepression of ERVs and induction of ISGs. We found that depletion of KAP1 alone leads not only to ERV reactivation but also to activation of the interferon-stimulated chemokines, CCL5 and CXCL10 in HeLa cells but not in PBMCs (Figure 6a and Figure S5a-c). Importantly, these genes are not direct KAP1 binding targets (ENCODE). This suggests that the KAP1 pathway is necessary to prevent ERV immune activation in certain cell contexts, although it may exert part redundancy with other epigenetic factors. In contrast, 5-AZA treatment potently activated ISGs in HeLa cells, 293T cells and primary cells showing that in

both cancer and normal cells, cytosine methylation acts to prevent viral mimicry and immune activation (Figure 6b). Note that these ISG promoters are not subject to direct DNA methylation regulation (ENCODE RRBS data) and rather that ISG activation has been shown to be dependent on MAVS signaling[8, 9]. Importantly, 5-AZA treatment affects some of the same loci as KAP1-depletion including the specific copies of HERVK14C that are detected with our primers (Figure 6). No further effect was observed from combined KAP1-depletion and 5-AZA treatment (Figure S5d). Once ISGs are activated, negative regulation of the interferon response ensures that these genes are switched off in a timely manner to prevent chronic immune activation[39]. Interestingly, we found that 12 immune response genes (Spreadsheet 6) were downregulated in stable KAP1 knockout HeLa cell clones, suggesting that the reactivated ERVs in these cells had played a role in modulating immune genes. Overall, these results show that KAP1 and cytosine methylation repress some of the same retrotransposon targets to prevent innate immune activation. This is relevant to the mode of action of 5-AZA-based drugs, which are used in the clinic. Since KAP1 depletion alone is not sufficient for global ISG induction, additional epigenetic factors likely contribute to retrotransposon repression in differentiated cells.

Discussion

The present work highlights the previously unknown role for KAP1 in repressing ERVs and ZNFs (and particularly KZNFs) in differentiated human cells. KAP1 therefore represents one epigenetic pathway necessary for preventing autoimmunity that can be triggered by ERV-derived nucleic acids and proteins. KAP1 is also relevant to the

mode of action of 5-AZA-based drugs since KAP1 and cytosine methylation repress some of the same ERVs (see Figure 7 for a model). The fact that KAP1 co-regulates ERVs and KZNFs, the latter which themselves recruit KAP1 to ERV repetitive DNA sites suggests that this mechanism may serve to fine-tune KZNF expression and maintain ERV repression. KAP1 constraint of KZNF transcription has been proposed before from binding data[32, 33], a notion which we now support here with functional data. Note that KAP1 knockout in differentiated cells impacts on cellular genes only modestly compared to the more pronounced effects observed following KAP1 knockout in mouse or human ESCs[13, 28]. This may relate to the increased requirement for KAP1 in early development or an increased number of KAP1 binding sites and / or different configuration of chromatin in embryonic cells, where heterochromatin spreads from repeats to genes[13, 28, 40].

Consistent with the presence of KAP1 and silent chromatin marks at certain ERV loci in multiple cell types, our data also reveal that a cohort of KZNFs exert a diverse mRNA and protein expression profile. These factors may function to regulate heterochromatin at a subset of the genome and to restrict infections from exogenous retroviruses. This is supported by our reporter experiments that demonstrate that KAP1 can mediate sequence-specific repression against incoming retroviruses. The fact that KAP1-repression is modest in differentiated cells may relate to the low expression levels of KZNFs. Intriguingly, the mouse KZFP, Zfp809 is rapidly degraded at the protein level in differentiated cells, restricting its potency mainly to embryonic cells [41]. An open question is whether KZNFs could become augmented in their expression in response to viral infections, interferon treatment

or other stimuli to make them more potent in differentiated cells. Finally, it's possible that KAP1 repression of common loci is dynamic and KAP1-bound ERV regulatory sequences may have been co-opted to activate cellular gene networks or downstream processes in specific contexts, a hypothesis worth exploring in the future. Interestingly, we found here that KAP1 regulates HERV-T and HERVK14C, both of which show evidence of co-option[42, 43]. KAP1 is often enriched at regulatory hubs together with other transcription factors (Figure S2), suggesting that these loci are active under specific conditions and once activated, the ERV-derived nucleic acids may help to trigger an immune response. Intriguingly, ERVs have been proposed to play a role in stimulating T-independent B cell responses[44].

Like KAP1 depletion, treatment of differentiated cells with 5-AZA results in the reactivation of KAP1-regulated retrotransposons including HERVK14C and to a lesser extent SVA D elements. However, depletion of KAP1 alone is not sufficient to broadly activate ISGs presumably because the effect on immuno-stimulatory retrotransposons is only modest. We propose that KAP1 exerts partial redundancy with other epigenetic pathways and that the level of redundancy at individual ERV integrants may relate to multiple parameters including their evolutionary age, local chromatin environment and nuclear position[14, 45, 46]. Future work will be necessary to determine exactly which factors act in addition to KAP1 to maintain ERV silencing in differentiated cells. Many epigenetic regulators exert cross-talk with DNA methylation including SETDB1[47], which can act independently or in concert with KAP1[48]. Other questions are the identities of the immuno-stimulatory retrotransposons and the nature of the signaling pathways they activate *in vivo*.

ERVs with bidirectional promoters can generate double stranded RNA recognized by MDA-5[8, 9] but we show that KAP1 and 5-AZA also repress other classes of retrotransposons in differentiated cells such as SVAs, and these may produce more potent pathogen associated molecular patterns (PAMPs).

Hundreds of KZNFs have recently been described to target specific sequences within repetitive DNA, including ERVs[17-19]. The relevance of such interactions, however, remains unknown. Our work illustrates that cell lines rather than ESCs could be used to determine the individual functions of a core set of KZNFs that are widely expressed through their targeted gene knockout. This would serve as an economical and easy screening method to assess the initial effects of KZNF knockout on the human genome before following up on promising candidates further *in vivo*. Most interestingly, KAP1 and core KZNFs in combination with other epigenetic factors may become new drug targets for cancer with the aim to reactivate subsets of ERVs and other retrotransposons and harness their natural ability to trigger innate and adaptive anti-tumour immunity.

Methods

Cell culture

Human teratoma-derived NTERA-2 cells (kind gift from Peter Andrews, University of Sheffield) were cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco) high glucose, supplemented with 2mM L-Glutamine, 10% fetal calf serum (FCS) and 1% Penicillin/Streptomycin (P/S). They were split 1:2 or 1:3 every 3-4 days by cell scraping. HEK293T (293T) and HeLa cells were grown in standard DMEM + 10% FCS and P/S and split 1:4 every 2 days using trypsin. Human primary CD4⁺ T cells were grown in Roswell Park Memorial Institute medium (RPMI, Gibco) supplemented with 10% human serum and 10 U/ml of recombinant IL-2 and activated using α CD3 and α CD28 pre-coated flasks for 72 hours during which time IL-2 was increased to 25U/ml. Peripheral blood mononuclear cells (PBMCs) were grown in RPMI supplemented with 20% FCS and 10U/ml of IL-2 and activated using 3 μ g/ml of Phytohemagglutinin-M (PHA) for 72 hours.

KAP1 knockout through CRISPR/Cas9 genome editing

Guide RNAs (sgRNAs) specific to several different exons were designed using the website: <http://crispr.mit.edu/> and cloned into the PX459 plasmid (Addgene), which was then transfected into HeLa and HEK293 T cells. After 24 hours, the cells were subjected to puromycin selection for 24 hours or until control cells had completely died. The bulk population was then used for single cell cloning. Knockout was assessed across a panel of clones by Western Blotting using the KAP1 antibody, MAB3662 (Millipore) and validated functionally using KAP1-KZNF reporter assays.

Primary cell isolation

PBMCs were isolated from a buffy coat or from fresh blood from healthy donors using lymphoprep (Axis-Shield). CD4⁺ T cells were obtained from the isolated PBMCs using the CD4⁺ T cell Isolation Kit (Miltenyl Biotec) according to manufacturer's instructions. The purity of the cells was verified by flow cytometry following antibody staining. Antibodies used are shown in Table 1.

Western blotting

1-2 x 10⁶ cells were washed with PBS and lysed in NuPAGE LDS sample buffer (Thermo Fisher) with 5% β-mercaptoethanol. Samples were sonicated for 90 seconds at 20Hz and heated at 95° C for 5 minutes. Lysates were then loaded onto handcast 10% SDS-polyacrylamide gels in tris/glycine/SDS buffer and mini-PROTEAN tanks (Biorad), followed by wet transfers onto Polyvinylidene Difluoride (PVDF) membranes. Antibodies used for blotting the membrane are listed in Table 1. All secondary antibodies were horseradish peroxidase-conjugated (GE healthcare) and membranes were developed using ECL kits (ECL, Prime or Select kits from Amersham).

Actin	Chemicon, MAB1501	mouse monoclonal
KAP1	Millipore, MAB3662	mouse monoclonal

PCNA	Millipore, clone PC10	mouse monoclonal
OCT3/4	Santa Cruz, sc-5279	mouse monoclonal
PE-Cy 5 CD3 (ε-chain)	BD Bioscience, 555341	mouse
FITC - CD4	BD Bioscience, 345768	mouse
PE - OCT3/4	eBioscience, 12-5841-80	rat
FITC IgG1 (κ Isotype Control)	BD Bioscience, 555909	mouse
PE-Cy 5 IgG2a (κ Isotype Control)	BD Bioscience, 555575	mouse
PE IgG2a (κ Isotype Control)	eBioscience, 12-4321-80	rat

Table 1. List of antibodies used.

shRNA lentiviral vectors and transduction

Hairpin sequences against human mRNAs were designed using the Clontech RNAi designer website (<http://bioinfo.clontech.com/rnaidesigner/>) and annealed into oligo duplexes. The duplexes were then cloned into HIV SIREN into *BamHI-EcoRI* sites and the products were checked by sequencing. VSV-G-pseudotyped lentiviral vectors were produced by co-transfecting 293T cells in 10cm plates with 1.5µg of the shRNA

plasmid, 1µg p8.91 and 1µg pMDG2 encoding VSV-G. The supernatant was harvested on two consecutive days (from 48h post-transfection) and used neat or concentrated via ultracentrifugation (20,000 × g for 2h at 4°C). Two days post transduction, the cells were selected with puromycin (2.5ug/ml) for 48 hours (or until control cells had completely died) prior to experiments.

RNA extraction and quantification

Total RNA was extracted using an RNeasy mini kit (Qiagen) and DNase (Ambion AM1907) treated. 500ng of RNA was used for cDNA production using SuperScript II Reverse Transcriptase (ThermoFisher) and random primers following the manufacturer's instructions. mRNA expression levels were quantified using quantitative reverse transcription PCR (qRT-PCR) using an ABI 7500 Real Time PCR System (Applied Biosystems). SYBR green Fast PCR mastermix (Life Technologies) was used. CT values for the test genes were normalized against those of *Gapdh* or *B2M* using the $-\Delta\Delta C_t$ method to calculate fold change. See Table 1 for primer sequences. Data are shown as means +/- standard deviation with the number of replicates stated in figure legends.

DNA methylation analysis

DNA was harvested using a DNeasy Blood & Tissue Kit (Qiagen) and 1µg of DNA was used for bisulfite conversion using an EpiTect Bisulfite Kit (Qiagen) following the manufacturer's protocol. 4µl of converted DNA was then amplified through PCR using the primer pairs described in the supplementary primer table (Table 1). Primers were designed using the site: <http://urogene.org/methprimer/> and the PCR

products were cloned using the TOPO TA-Cloning Kit (Thermo Fisher Scientific) and the T7P primer was used to sequence the products. DNA methylation status of the TOPO clones was measured using the QUMA online tool (<http://quma.cdb.riken.jp>) by the Riken Institute.

GFP reporter assay

The HERVK14C LTR was cloned into a PGK-GFP plasmid in place of the PGK promoter at XhoI-BamHI sites. The consensus lysine PBS sequence was cloned upstream of the LTR via an oligo-duplex strategy into the XhoI site while the Chromosome 15 HERVK14C specific lysine PBS was cloned into the backbone through a PCR strategy into the XhoI-BamHI sites. The final products were verified via sequencing.

Cells were plated at a concentration of 10^5 cells/ml in 24 well plates. After 6 hours, wells were transduced with VSV-pseudotyped GFP vectors at increasing doses (normalized between vectors by the number of transducing units, which were calculated by first titering vectors on kap1 knockout 293T cells) and fresh media was replenished after 24 hours. After a further 48 hours, the cells were fixed in 1% PFA, washed in PBS and GFP was read using flow cytometry.

Intracellular Oct4 staining

1×10^6 cells per condition were fixed and permeabilized using intracellular staining buffers. The cells were then stained with Oct4-PE or isotype control antibodies (see Table 1), washed and analyzed by flow cytometry.

Luciferase reporter assay

Luciferase assays were conducted according to the Promega dual-Luciferase kit instructions. 293T cells were plated 6 hours before transfection at a concentration of 10^5 cells/ml in 24 well plates. Following ratios defined before[14], 200 ng of KRAB ZNF plasmid DNA, 20 ng of Luciferase reporter plasmid DNA and 2 ng of pRT-TK_Renilla control plasmid were co-transfected (10:1:0.1 ratio) using 1.5 μ l of Eugene 6 (Promega) and 30 μ l of Opti-MEM (Gibco). 48 hours post-transfection, cells were lysed and luciferase was measured using the Dual Luciferase[®] assay kit (Promega, E1910) in an opti-plate using a Glomax 96 microplate Luminometer's (Promega) Dual Glow program. The Renilla-encoding plasmid was used as an internal control for transfection efficiency normalization. Firefly to Renilla ratio was then further normalized against the empty vector or negative control as repression readouts and expressed in percentages where the control is set to 100%.

mRNA-sequencing and analysis

Total RNA Samples were processed using Illuminas TruSeq Stranded mRNA LT sample preparation kit (RS-122-2101) according to manufacturer's instructions with some deviations: Libraries to be multiplexed in the same run were pooled in equimolar quantities, calculated from Qubit and Bioanalyser fragment analysis. Samples were sequenced on a NextSeq 500 instrument (Illumina, San Diego, US) using a 43bp paired end run resulting in >15million reads per sample. Run data were demultiplexed and converted to fastq files using Illumina's bcl2fastq Conversion Software v2.16. Fastq files were then aligned to the human genome NCBI build 37.2 using Tophat 2.014 then deduplicated using Picard Tools 1.79. Reads per transcript were counted using HTSeq and differential expression was estimated using the

BioConductor package DESeq2. P-values were adjusted for multiple testing with the Benjamini-Hochberg false-discovery-rate (FDR) procedure. Genes were considered up or downregulated where differential affects were > 2 fold and where adjusted p values were <0.05. For analysis of repeats, TrimGalore v0.4.0 was used to remove adaptors and trim read ends using default parameters. Reads were mapped against the RepBase v20.06 human library using Bowtie2v2.2.4. The samtools v.1.19 idxstat utility was used to extract the number of mapped reads per repeat, that were inputted to DESeq2 to identify differentially expressed repeats. P-values were adjusted for multiple testing with the Benjamini-Hochberg false-discovery-rate (FDR) procedure. Gene ontology analysis was conducted using the DAVID website: <https://david-d.ncifcrf.gov/>[49, 50]. Public mRNA-sequencing data was used to determine if KZNFs were expressed in different cell types (see data access for accession numbers). A gene was considered to be expressed if it had an rpkm value of >0.5. The top 100 KZNFs expressed in 293T cells were determined by sorting mRNA-sequencing data on rpkm values.

ChIP-sequencing analysis

Chromatin-immunoprecipitation-sequencing data for human ESCs and 293T cells were downloaded from NCBI Gene Expression Omnibus (GEO) under accession numbers GSE57989 (HuESC) and GSE27929 (293T). TrimGalore v0.4.0 was used to remove adaptors and trim read ends, and reads were mapped against the human genome (hg19 assembly) using Bowtie2 v2.2.4. Peaks were called in each replicate using Macs2 v 2.1.1, and the bioconductor package DiffBind (<https://bioconductor.org/packages/DiffBind>) was used to construct Venn diagrams

and identify overlapping ESC and 293T peaks. Human repeat and gene locations were downloaded from the UCSC browser (RepeatMasker and RefGene tables) and the repeats and genes closest to the overlapping ESC/293T peaks were identified using bedtools-2.17.0. Chip-sequencing correlations were analysed using the Chip-Cor website: http://ccg.vital-it.ch/chipseq/chip_cor.php.

Statistical analysis

All data in the figures are presented as the standard deviation (where there are three or more samples) assessed by using two tailed, unpaired Student t tests (see figure legends for details). A P-value of < 0.05 was considered statistically significant (***= $p < 0.001$, **= $p < 0.01$ and *= $p < 0.05$).

Ethics statement

Healthy adult blood donors provided written informed consent. Culture of primary peripheral blood mononuclear cells from blood donors has been reviewed and granted ethical permission by the National Research Ethics Service through The Joint UCL/UCLH Committees on the Ethics of Human Research (Committee Alpha) 2 December 2009; reference number 06/q0502/92.

Data access

mRNA-sequencing data are being submitted to the NCBI Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) database and other data are included in the article and its supplementary information files. Accession numbers for the public data are as follows: 293T: GSE27929 (Chip-seq), HuESC: GSE57989 (Chip-seq), 293T:

GSE44267 (mRNA-seq), HeLa: this study (mRNA-seq), Macrophages: GSE36952 (mRNA-seq), CD4+ T cells: GSE69549 (mRNA-seq), HuESC (mRNA-seq).

Acknowledgements

We thank Peter Andrews for the NTERA-2 cells, Didier Trono for reagents previously generated in his lab and David Haussler for the SVA, LINE1, ZNF91 and ZNF93 constructs. We thank Connor Husovsky for technical assistance and Steen Ooi, Pierre Maillard and members of UCL Cruciform wing 3.3 for advice. We also thank the UCL Genomics Unit and Tony Brooks for carrying out the mRNA-sequencing library preparation, sequencing and initial data analysis. This work was supported through a UCL Grand Challenges Studentship awarded to CHCT (code 518099) and a Sir Henry Dale Fellowship, which is jointly funded by the Wellcome Trust and Royal Society (Grant number 101200/Z/13/Z) awarded to HMR. RG is funded by the MRC (MC_UU_12014/10 and GJT and MTR by the Wellcome Trust (108183) and ERC (339223) respectively.

Author contributions

CHCT conceived, designed and performed experiments, analyzed the data and wrote the paper. LRM, MTR, RG and GT contributed to reagents, experiments, and ideas. LC and JH performed bioinformatics analyses. HMR conceived the study, analyzed data and wrote the paper. All authors read and approved the final manuscript.

Conflict of interest

The authors declare that they have no conflicts of interests.

Figure Legends

Figure 1. HERVK14C is repressed by KAP1 in undifferentiated and differentiated cells

(A) Schematic diagram showing the age of HERVK14C. (B) Chromosome map showing the HERVK14C loci with at least one internal open reading frame and long terminal repeats. Note that many copies are present on the Y chromosome so are not relevant in female cells. (C) Venn diagram showing the chromosome copies detected by the two primer sets. (D) qRT-PCR expression of endogenous repeats following shRNA-mediated KAP1 depletion in NTERA-2 cells. Results were normalized to $\beta 2$ microglobulin (B2M). KAP1 expression levels were verified by qRT-PCR and Western blot. A representative experiment of 2 experiments is shown. Two-tailed unpaired t tests were done (HERVK14C_2 p-value = 0.002300). (E and F) qRT-PCR expression of endogenous repeats following KAP1 knockout in HeLa (E) and 293T cells (F). Results were normalized to $\beta 2$ microglobulin (B2M). KAP1 expression levels were verified by qRT-PCR and Western blot. Two-tailed unpaired t tests were done and p-values are: (E) HERVK14C_1: 0.0017, HERVK14C_2: 0.0044, SVA D VNTR: 0.0243 (F) HERVK14C_2: <0.0001. Clones 8, 12 and 15 (from E) were selected for mRNA-sequencing.

Figure 2. A common role for KAP1 in repressing ERVs and ZNFs

(A) Boxplots showing repeats significantly upregulated (>2 fold where $p < 0.05$) in knockout compared to wildtype HeLa cells based on mRNA-sequencing data. HERV-T and HERV-S but not HERVK14C also reached significance when only adjusted p-

values were considered, where differences are compared to the whole of repbase. KAP1 binds to these ERVs (ENCODE data) and binding data is shown in Figure S2a. (B) Evolutionary tree showing the age of KAP1-repressed HERVs identified in A). Here, HERV-K refers to HERVK14C. Estimated ages of stated lineages are given and marked with a star. Myr: million years (C) The 38 upregulated genes identified (>2 fold where $p_{adj} < 0.05$) were converted to DAVID IDs and used for gene ontology analysis. 3 gene clusters were significantly enriched (p -value = < 0.05 , drawn on the plot as a dotted line) in the data set: zinc finger proteins ($p = 2.2 \times 10^{-7}$), negative regulators of metabolic processes ($p = 0.0036$) and angiogenesis ($p = 0.011$). Venn diagrams on the right show numbers of upregulated genes, ZNFs and KZNFs and the overlap. (D) The nature of conserved KAP1 binding sites between human ESCs[13] and 293Ts (ENCODE) (see Figure S3d) is shown (left pie chart). ERVs are enriched compared to their relative abundance in the genome (right pie chart). (E) The 614 KAP1 common binding sites (see Figure S3d) were interrogated for their nearest gene and this gene list was converted to DAVID IDs and used for gene ontology analysis. 4 gene clusters were significantly enriched (p -value < 0.05 , drawn on the plot as a dotted line): zinc finger proteins ($p = 1.1 \times 10^{-19}$), fibronectin folding ($p = 0.016$), protein phosphatase ($p = 0.033$), and synapse ($p = 0.047$). Venn diagrams on the right show numbers of KAP1-bound loci, ZNFs and KZNFs and the overlap. (F) Comparison of the distribution of KAP1-bound ERVs and LINE1 elements between undifferentiated cells (human ESCs), differentiated cells (293Ts) and KAP1 common sites (sites conserved between human ESCs and 293Ts).

Figure 3. KAP1 common binding sites are enriched for repressive histone marks in

multiple cell types and overlap KZNF ERV targets

(A) Genomic coordinates of the common KAP1 sites identified in Figure S3D (conserved between ESCs and 293Ts) were subjected to ChIP-seq correlation analyses using ChIP-Cor software (see methods). Each plot shows duplicate ChIP-seq experiments from ENCODE. (B) The binding sites (where known[17, 19]) within repeats and genes of the top 100 most highly expressed (defined by RNAseq rpkm) KZNFs in 293T cells. These KZNFs were also confirmed to be expressed at the protein level (<http://www.proteinatlas.org/cell>). (C) The binding sites within repeats and genes of KAP1 at common KAP1 sites. (D) KZNFs within the top 100 most highly expressed group that are known to bind ERVs were assessed for their age and type of ERV that they bind to.

Figure 4. KAP1 restricts incoming ERVs through the primer binding site sequence

(A) Schematic diagram showing reporter constructs used. The 18-bp PBS sequence was cloned upstream of the HERVK14C LTR promoter (which is identical in all vectors) in an antisense orientation. (B) NTERA-2 cells were transduced with either an empty vector (WT) or the same vector containing an shRNA against KAP1 (KD) prior to puromycin selection and transduction with increasing doses of GFP reporter vectors. GFP expression was analysed 72 hours post reporter transduction. A representative experiment of 2 experiments is shown. Two-tailed unpaired t test p-values: PBS-LTR-GFP: 0.0148; PBSChr15-LTR-GFP: 0.0377 (C) The same as (B) except that here KAP1 wildtype (WT) and knockout (KO) 293T cells were used. P-values: PBS-LTR-GFP: 0.0044; PBSChr15-LTR-GFP: 0.0077.

Figure 5. PBS-dependent *de novo* repression of ERVs is conserved in primary cells

(A) qRT-PCR expression of endogenous repeats following shRNA-mediated KAP1 depletion in PBMCs. Results were normalized to $\beta 2$ microglobulin (B2M). (B) PBMCs were transduced with increasing doses of GFP reporter vectors (vectors were normalized to the same number of infectious units after titering them on permissive cells), left plot. PBS-repression was verified to be not due to lack of integration as vectors were integrated at similar relative levels as measured by GFP Taqman qPCR, right plot. A 293T cell line with a single vector copy integrant (PGK-GFP) was used as a control to estimate the absolute copy numbers. A representative experiment of 3 experiments is shown here. Two-tailed unpaired t test p-value 0.0288 (dose 3). (C) The expression of 340 KZNFs was assessed in multiple cell types using public mRNAsequencing data and venn diagrams show the overlap in expression profiles. (D) Western blots showing the protein expression levels of ZNF37A, ZNF33A, and ZNF320 in HeLa cells, 293T cells, and NTERA-2 cells.

Figure 6. Cytosine methylation acts on KAP1-regulated retrotransposons and prevents viral mimicry and innate immune activation

(A) qRT-PCR expression of endogenous repeats (left) and ISGs (right) following shRNA-mediated KAP1 depletion in HeLa cells. n=2 and two-tailed unpaired t test p-values are: HERVK14C: 0.025, 0.003; SVA VNTR: 0.00027, 0.01; ISG56: 0.002; CCL5: 0.007, 0.00003; CXCL10: 0.0008; Mxa: 0.011; IKba: 0.008, 0.01. See also Figure S5 for ISG expression in KAP1-depleted PBMCs. (B) qRT-PCR expression of endogenous repeats (left) and ISGs (right) following 5-AZA treatment of 293T cells, HeLa cells and

PBMCs. A representative experiment of two is shown in each case. Two-tailed unpaired t tests were performed. 293T cell p-values: HERVK14C: 0.018; SVA D VNTR: 0.005; CCL5: 0.004; CXCL10: 0.003; Mxa: 0.00006; IKba: 0.007. HeLa cell p-values: ISG56: 0.005; CCL5: 0.0006; CXCL10: 0.0007; Mxa: 0.0001; IFNb: 0.02. PBMC p-values: HERVK14C: 0.1; SVA D VNTR: 0.003; ISG56: 0.001; CCL5: 0.012; CXCL10: 0.012; Mxa: 0.02; IKba: 0.08. Results were normalized to Gapdh and B2M (Gapdh results shown).

Figure 7: **Model**

KAP1 controls ERVs and ZNFs in differentiated human cells. KAP1 and SETDB1 binding are detected at ERVs and ZNFs in differentiated cells and overlap with the silent chromatin mark H3K9me3 as well as cytosine methylation, which we detect at HERVK14C and SVAs. The KAP1-KZNF pathway is functionally intact in differentiated cells including PBMCs. 5-AZA treatment leads to the inhibition of cytosine methylation at KAP1-regulated retrotransposons including ERVs, inducing reactivation of ERV and SVA transcription. These retrotransposons induce viral mimicry because they can produce double stranded RNA, and ERVs can potentially also produce DNA through reverse transcription. Viral mimicry results in nucleic acid sensing and induction of ISGs. KAP1 depletion is not sufficient for global ISG induction suggesting it exerts partial redundancy at repressing ERVs with other epigenetic regulators in differentiated human cells.

Figure S1. **Validation of cell types and functional validation of KAP1 knockout.**

(A) NTERA-2 and 293T cells were stained for Oct4 intracellularly and were subjected to flow cytometry analysis. A PE isotype control antibody was used as a negative control. (B) KAP1 knockout HeLa cells were validated using a known KAP1-KZFP target sequence[15, 16]: Wildtype HeLa cells, KAP1 knockout cells and KAP1 reconstituted knockout cells (KAP1 expression levels are shown by Western blot) were doubly transduced with a PBS-Pro-GFP reporter vector (from [12]) and a vector expressing either the cognate ZFP, Zfp809 or a control ZFP, Zfp819. In MOCK controls, the ZFP-expression vectors were replaced with same volume of media. Two-tailed unpaired t test p-values are: HeLa wildtype: 0.033; KAP1 KO clone_1: 0.3185; KAP1 KO clone_2: 0.0036. (C) KAP1 knockout 293T cells were validated using known KAP1-KZNF target sequences (constructs were a kind gift from David Haussler) [14]. KAP1 wildtype and knockout 293T cells were co-transfected with the luciferase reporter construct, a renilla control plasmid and constructs expressing either ZNF91 or ZNF93 (see methods). Two-tailed unpaired t test p-values are: WT SVA_ZNF91: <0.0001; WT LINE1_ZNF93: 0.0001; KO SVA_ZNF91: 0.6008; KO LINE1_ZNF93: 0.0841. (D) DNA methylation analysis of endogenous SVA D VNTRs in 293T cells. Primers were designed on a SVA D VNTR copy on chromosome 7 but primers recognize 219 copies of SVAs, some of which exhibit CpG deletions or mutations (shown by “x” on the CpG map). PCR for the Oct4 gene body was used as a positive endogenous control for cytosine methylation.

Figure S2. Examples of KAP1-regulated repeat elements.

(A) Examples of ERV loci that are bound by KAP1 and potentially act as regulatory hubs through multiple transcription factor interactions as revealed by ENCODE data.

(B) Repeats significantly downregulated in KAP1 knockout HeLa clones compared to control cells based on mRNA sequencing data.

Figure S3. Cellular genes are among the targets of KAP1 regulation.

(A) Examples of cellular genes upregulated or downregulated upon KAP1 knockout in HeLa cells based on mRNA-sequencing data. BigWig files were uploaded to the UCSC genome browser as URL links and the same scale applied to all samples (B) The top 100 most significantly downregulated genes (>2 fold with a padj value of <0.05) were converted to DAVID IDs and used for gene ontology analysis. All shown gene clusters were significantly enriched (p-values = <0.05) in the data set: signal peptide (p=8.9 x 10⁻¹⁹), glycoprotein (p=3.6 x 10⁻¹⁹), cell adhesion (p=2.5x 10⁻¹⁰), extracellular matrix (p=1.6x 10⁻⁸), EGF-like region (p=0.000004), wounding response (p=0.000006), plasma membrane (p=0.000029), cell-cell adhesion (p=0.000160), and innate immune response (p=0.000780). (C) Table showing that all upregulated KZNFs (from Figure 2c) are KAP1-bound according to ENCODE data. (D) Venn diagrams showing that some KAP1 binding sites are common between human ESCs and 293T cells. 614 common binding sites were identified based on their presence in duplicate KAP1 ChIP-sequencing experiments in human ESCs as well as in duplicate KAP1 ChIP-sequencing experiments in 293T cells. Data are from [13] and ENCODE (293T cells).

Figure S4. A role for KAP1 in repression of ERVs in primary cells

(A) Shows an independent PBMC experiment with a different donor to that shown in Figure 5a. (B and C) Primary CD4+ T cells were activated and transduced with KAP1 knockdown or control vectors. Cells were harvested at different time points for (B)

Western blot and (C) qRT-PCR. (D and E) DNA methylation status of the HERVK14C LTR region on Chromosome 15 in CD4+ T cells (D) and KAP1 wildtype and knockout Hela cells (E) as tested using bisulphite sequencing.

Figure S5. KAP1 depletion is not sufficient for global ISG induction

(A and B) PBMCs depleted for KAP1 as shown in Figure 5a and Figure S4a were assessed for induction of ISGs by qRT-PCR. (C) A repeat independent experiment to that shown in Figure 6a is shown here. Two-tailed unpaired t tests were performed and p-values are: L1PA4: 0.009, 0.003; HERVK14C: 0.006, 0.004; SVA D VNTR: 0.012; CCL5: 0.005, 0.007. (D) qRTPCR expression of endogenous repeats (left) and ISGs (right) following shRNA-mediated KAP1 depletion or 5-AZA treatment or a combination of both in 293T cells, n=2. Two-tailed unpaired t test p-values are: HERVK14C: 0.00012, 0.0007; SVA VNTR: 0.0017; ISG56: 0.0016, 0.02, 0.00097; CCL5: 0.0003, 0.00006; CXCL10: 0.0001, 0.0007; MxA: 0.005, 0.0009, 0.00034; Ikba: 0.012, 0.0077.

References

1. Chuong EB, Elde NC, Feschotte C: **Regulatory activities of transposable elements: from conflicts to benefits.** *Nat Rev Genet* 2017, **18**(2):71-86.
2. de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD: **Repetitive elements may comprise over two-thirds of the human genome.** *PLoS Genet* 2011, **7**(12):e1002384.
3. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al*: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**(6822):860-921.
4. Babaian A, Mager DL: **Endogenous retroviral promoter exaptation in human cancer.** *Mobile DNA* 2016, **7**:24.
5. Chuong EB, Elde NC, Feschotte C: **Regulatory evolution of innate immunity through co-option of endogenous retroviruses.** *Science* 2016, **351**(6277):1083-1087.
6. Robbez-Masson L, Rowe HM: **Retrotransposons shape species-specific embryonic stem cell gene expression.** *Retrovirology* 2015, **12**:45.
7. Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A: **Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice.** *Oncogene* 2008, **27**(3):404-408.
8. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, Hein A, Rote NS, Cope LM, Snyder A *et al*: **Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses.** *Cell* 2015, **162**(5):974-986.
9. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, Han H, Liang G, Jones PA, Pugh TJ *et al*: **DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts.** *Cell* 2015, **162**(5):961-973.
10. Antony JM, van Marle G, Opii W, Butterfield DA, Mallet F, Yong VW, Wallace JL, Deacon RM, Warren K, Power C: **Human endogenous retrovirus glycoprotein-mediated induction of redox reactants causes oligodendrocyte death and demyelination.** *Nature neuroscience* 2004, **7**(10):1088-1095.
11. Walter M, Teissandier A, Perez-Palacios R, Bourc'his D: **An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells.** *Elife* 2016, **5**.
12. Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J *et al*: **KAP1 controls endogenous retroviruses in embryonic stem cells.** *Nature* 2010, **463**(7278):237-240.
13. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, Tieng V, Quenneville S, Trono D: **Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements.** *Genome Res* 2014, **24**(8):1260-1270.
14. Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, Paten B, Salama SR, Haussler D: **An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons.** *Nature* 2014, **516**(7530):242-245.

15. Wolf D, Goff SP: **TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells.** *Cell* 2007, **131**(1):46-57.
16. Wolf D, Goff SP: **Embryonic stem cells use ZFP809 to silence retroviral DNAs.** *Nature* 2009, **458**(7242):1201-1204.
17. Imbeault M, Helleboid PY, Trono D: **KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks.** *Nature* 2017.
18. Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A, Albu M, Weirauch MT, Radovani E, Kim PM *et al*: **C2H2 zinc finger proteins greatly expand the human regulatory lexicon.** *Nat Biotechnol* 2015, **33**(5):555-562.
19. Schmitges FW, Radovani E, Najafabadi HS, Barazandeh M, Campitelli LF, Yin Y, Jolma A, Zhong G, Guo H, Kanagalingam T *et al*: **Multiparameter functional diversity of human C2H2 zinc finger proteins.** *Genome Res* 2016, **26**(12):1742-1752.
20. Friedman JR, Fredericks WJ, Jensen DE, Speicher DW, Huang XP, Neilson EG, Rauscher FJ, 3rd: **KAP-1, a novel corepressor for the highly conserved KRAB repression domain.** *Genes Dev* 1996, **10**(16):2067-2078.
21. Groner AC, Meylan S, Ciuffi A, Zangger N, Ambrosini G, Denervaud N, Bucher P, Trono D: **KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading.** *PLoS Genet* 2010, **6**(3):e1000869.
22. Iyengar S, Ivanov AV, Jin VX, Rauscher FJ, 3rd, Farnham PJ: **Functional analysis of KAP1 genomic recruitment.** *Mol Cell Biol* 2011, **31**(9):1833-1847.
23. Ivanov AV, Peng H, Yurchenko V, Yap KL, Negorev DG, Schultz DC, Psulkowski E, Fredericks WJ, White DE, Maul GG *et al*: **PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing.** *Mol Cell* 2007, **28**(5):823-837.
24. Matsui T, Leung D, Miyashita H, Maksakova IA, Miyachi H, Kimura H, Tachibana M, Lorincz MC, Shinkai Y: **Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET.** *Nature* 2010, **464**(7290):927-931.
25. Rowe HM, Friedli M, Offner S, Verp S, Mesnard D, Marquis J, Aktas T, Trono D: **De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET.** *Development* 2013, **140**(3):519-529.
26. Brattas PL, Jonsson ME, Fasching L, Nelander Wahlestedt J, Shahsavani M, Falk R, Falk A, Jern P, Parmar M, Jakobsson J: **TRIM28 Controls a Gene Regulatory Network Based on Endogenous Retroviruses in Human Neural Progenitor Cells.** *Cell Rep* 2017, **18**(1):1-11.
27. Ecco G, Cassano M, Kauzlaric A, Duc J, Coluccio A, Offner S, Imbeault M, Rowe HM, Turelli P, Trono D: **Transposable Elements and Their KRAB-ZFP Controllers Regulate Gene Expression in Adult Tissues.** *Dev Cell* 2016, **36**(6):611-623.
28. Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, Tarabay Y, Viville S, Jakobsson J, Pfaff SL, Trono D: **TRIM28 repression of**

- retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells.** *Genome Res* 2013, **23**(3):452-461.
29. Collins PL, Kyle KE, Egawa T, Shinkai Y, Oltz EM: **The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes.** *Proc Natl Acad Sci U S A* 2015, **112**(27):8367-8372.
 30. Sadic D, Schmidt K, Groh S, Kondofersky I, Ellwart J, Fuchs C, Theis FJ, Schotta G: **Atrx promotes heterochromatin formation at retrotransposons.** *EMBO Rep* 2015, **16**(7):836-850.
 31. Santoni de Sio FR, Massacand J, Barde I, Offner S, Corsinotti A, Kapopoulou A, Bojkowska K, Dagklis A, Fernandez M, Ghia P *et al*: **KAP1 regulates gene networks controlling mouse B-lymphoid cell differentiation and function.** *Blood* 2012, **119**(20):4675-4685.
 32. Fietze S, O'Geen H, Blahnik KR, Jin VX, Farnham PJ: **ZNF274 recruits the histone methyltransferase SETDB1 to the 3' ends of ZNF genes.** *PLoS One* 2010, **5**(12):e15082.
 33. O'Geen H, Squazzo SL, Iyengar S, Blahnik K, Rinn JL, Chang HY, Green R, Farnham PJ: **Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs.** *PLoS Genet* 2007, **3**(6):e89.
 34. Zuin J, Dixon JR, van der Reijden MI, Ye Z, Kolovos P, Brouwer RW, van de Corput MP, van de Werken HJ, Knoch TA, van IWF *et al*: **Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells.** *Proc Natl Acad Sci U S A* 2014, **111**(3):996-1001.
 35. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S *et al*: **Towards a knowledge-based Human Protein Atlas.** *Nat Biotechnol* 2010, **28**(12):1248-1250.
 36. Wolf D, Hug K, Goff SP: **TRIM28 mediates primer binding site-targeted silencing of Lys1,2 tRNA-utilizing retroviruses in embryonic cells.** *Proc Natl Acad Sci U S A* 2008, **105**(34):12521-12526.
 37. Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, Shinkai Y, Mager DL, Jones S, Hirst M *et al*: **DNA methylation and SETDB1/H3K9me3 regulate predominantly distinct sets of genes, retroelements, and chimeric transcripts in mESCs.** *Cell Stem Cell* 2011, **8**(6):676-687.
 38. Wolf G, Yang P, Fuchtbauer AC, Fuchtbauer EM, Silva AM, Park C, Wu W, Nielsen AL, Pedersen FS, Macfarlan TS: **The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses.** *Genes Dev* 2015, **29**(5):538-554.
 39. Taylor KE, Mossman KL: **Recent advances in understanding viral evasion of type I interferon.** *Immunology* 2013, **138**(3):190-197.
 40. Rebollo R, Karimi MM, Bilenky M, Gagnier L, Miceli-Royer K, Zhang Y, Goyal P, Keane TM, Jones S, Hirst M *et al*: **Retrotransposon-induced heterochromatin spreading in the mouse revealed by insertional polymorphisms.** *PLoS Genet* 2011, **7**(9):e1002301.
 41. Wang C, Goff SP: **Differential control of retrovirus silencing in embryonic cells by proteasomal regulation of the ZFP809 retroviral repressor.** *Proc Natl Acad Sci U S A* 2017, **114**(6):E922-E930.

42. Blanco-Melo D, Gifford RJ, Bieniasz PD: **Co-option of an endogenous retrovirus envelope for host defense in hominid ancestors.** *Elife* 2017, **6**.
43. Goke J, Lu X, Chan YS, Ng HH, Ly LH, Sachs F, Szczerbinska I: **Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells.** *Cell Stem Cell* 2015, **16**(2):135-141.
44. Zeng M, Hu Z, Shi X, Li X, Zhan X, Li XD, Wang J, Choi JH, Wang KW, Purrington T *et al*: **MAVS, cGAS, and endogenous retroviruses in T-independent B cell responses.** *Science* 2014, **346**(6216):1486-1492.
45. Barklis E, Mulligan RC, Jaenisch R: **Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells.** *Cell* 1986, **47**(3):391-399.
46. Reddy KL, Zullo JM, Bertolino E, Singh H: **Transcriptional repression mediated by repositioning of genes to the nuclear lamina.** *Nature* 2008, **452**(7184):243-247.
47. Leung D, Du T, Wagner U, Xie W, Lee AY, Goyal P, Li Y, Szulwach KE, Jin P, Lorincz MC *et al*: **Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1.** *Proc Natl Acad Sci U S A* 2014, **111**(18):6690-6695.
48. Yang BX, El Farran CA, Guo HC, Yu T, Fang HT, Wang HF, Schlesinger S, Seah YF, Goh GY, Neo SP *et al*: **Systematic identification of factors for provirus silencing in embryonic stem cells.** *Cell* 2015, **163**(1):230-245.
49. Huang da W, Sherman BT, Lempicki RA: **Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.** *Nat Protoc* 2009, **4**(1):44-57.
50. Huang da W, Sherman BT, Lempicki RA: **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.** *Nucleic Acids Res* 2009, **37**(1):1-13.

Figure 1

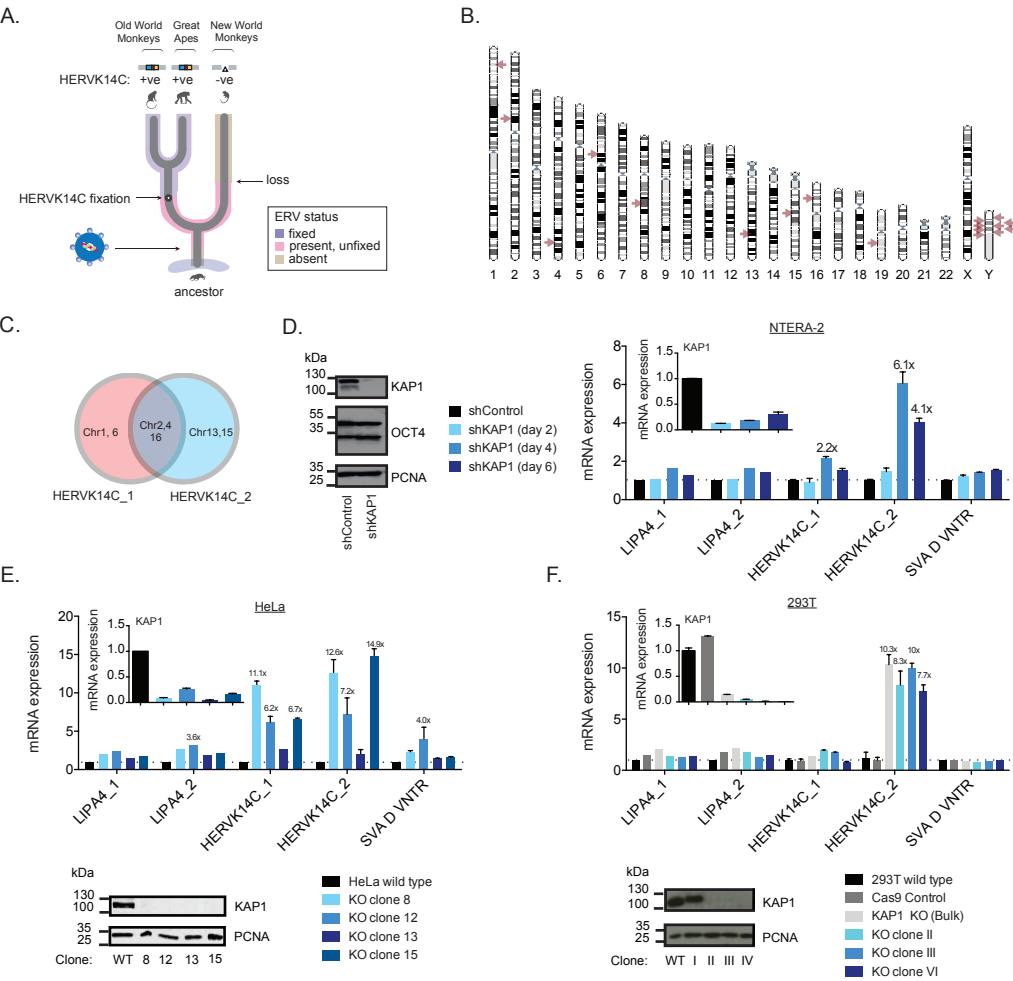


Figure 2

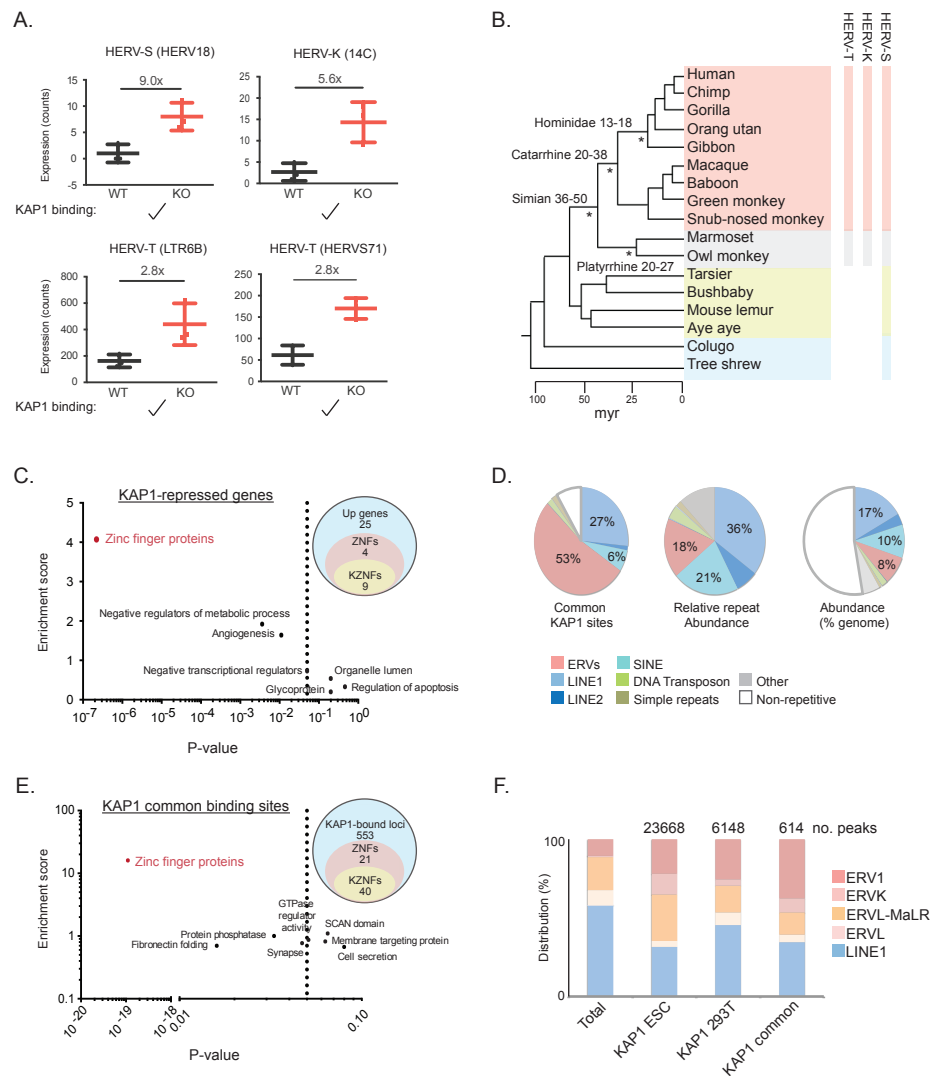


Figure 3

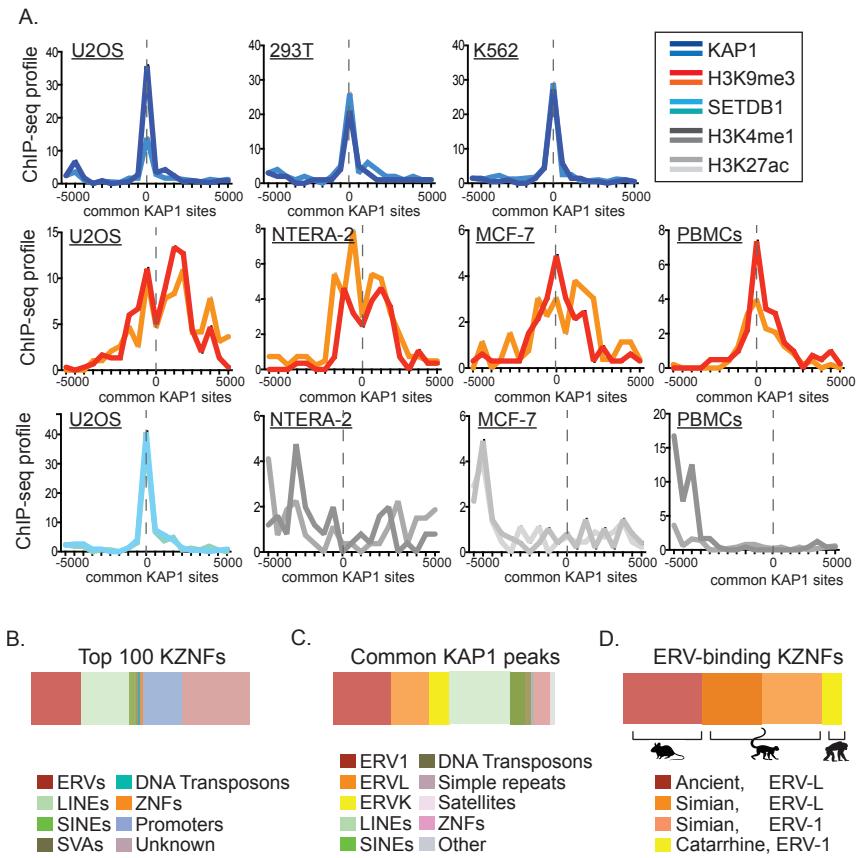


Figure 4

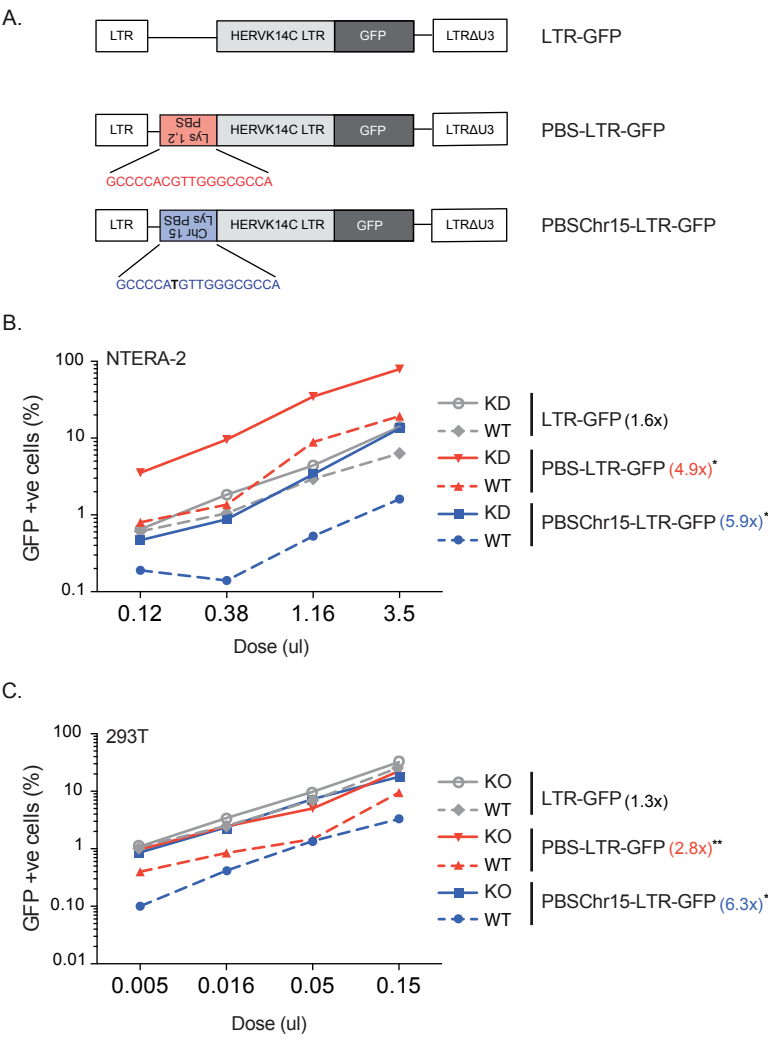


Figure 5

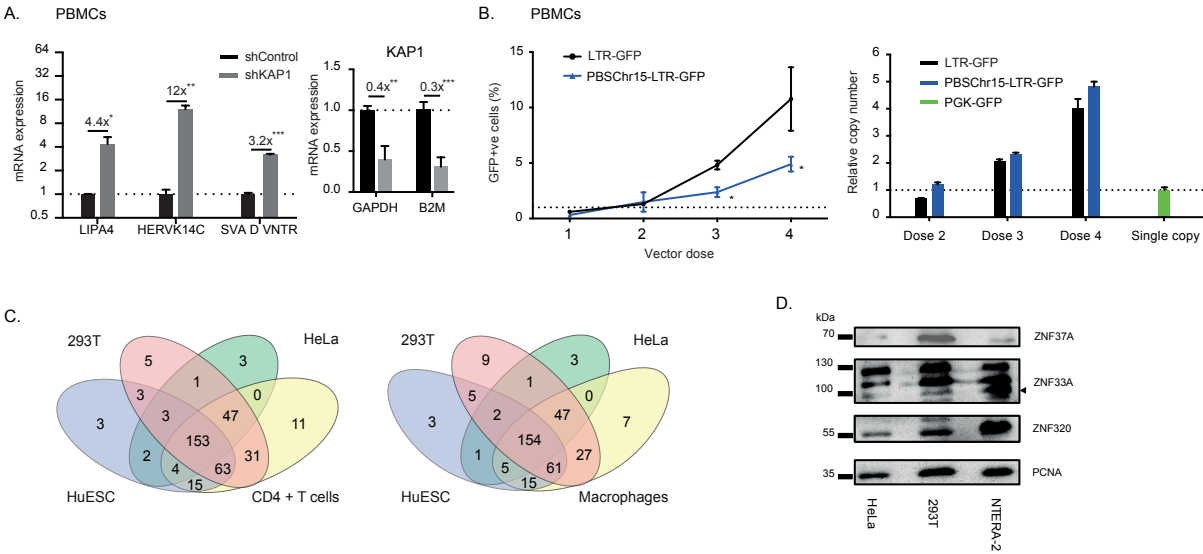


Figure 6

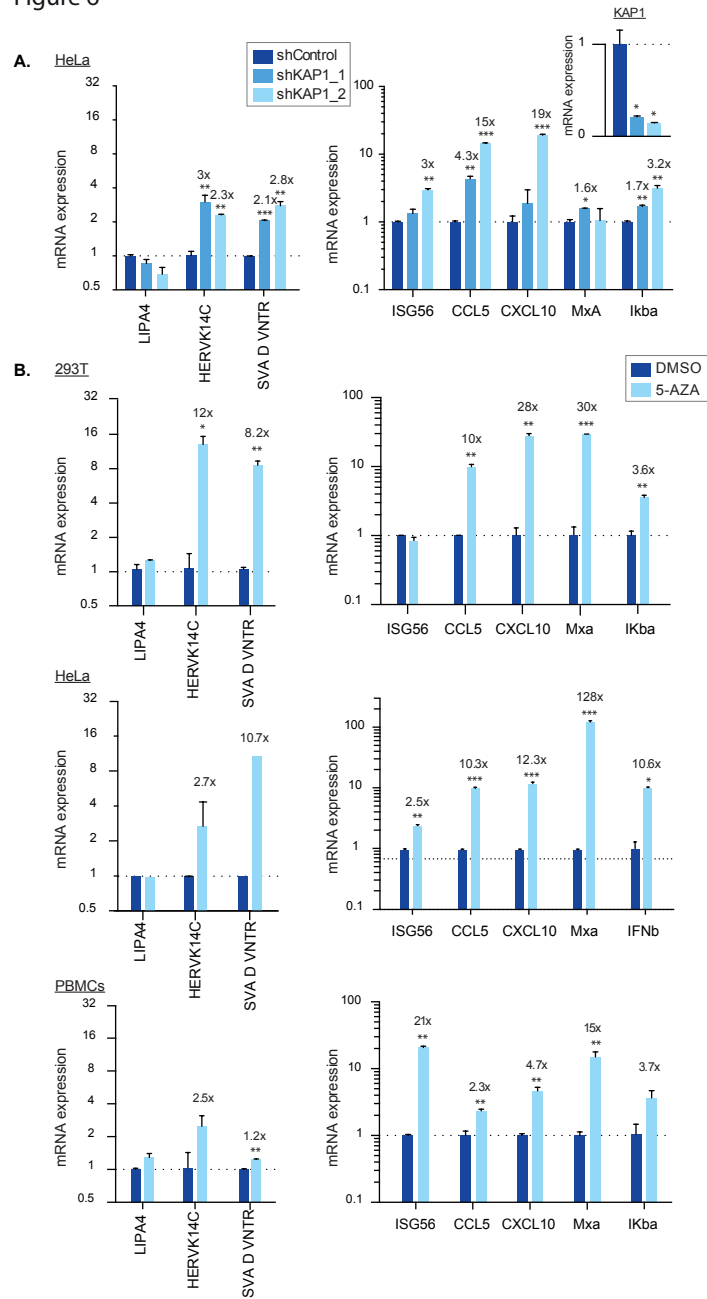


Figure 7

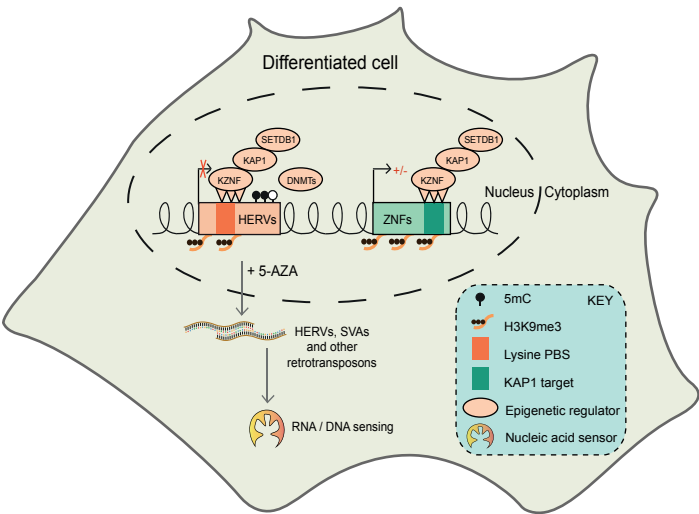


Figure S1

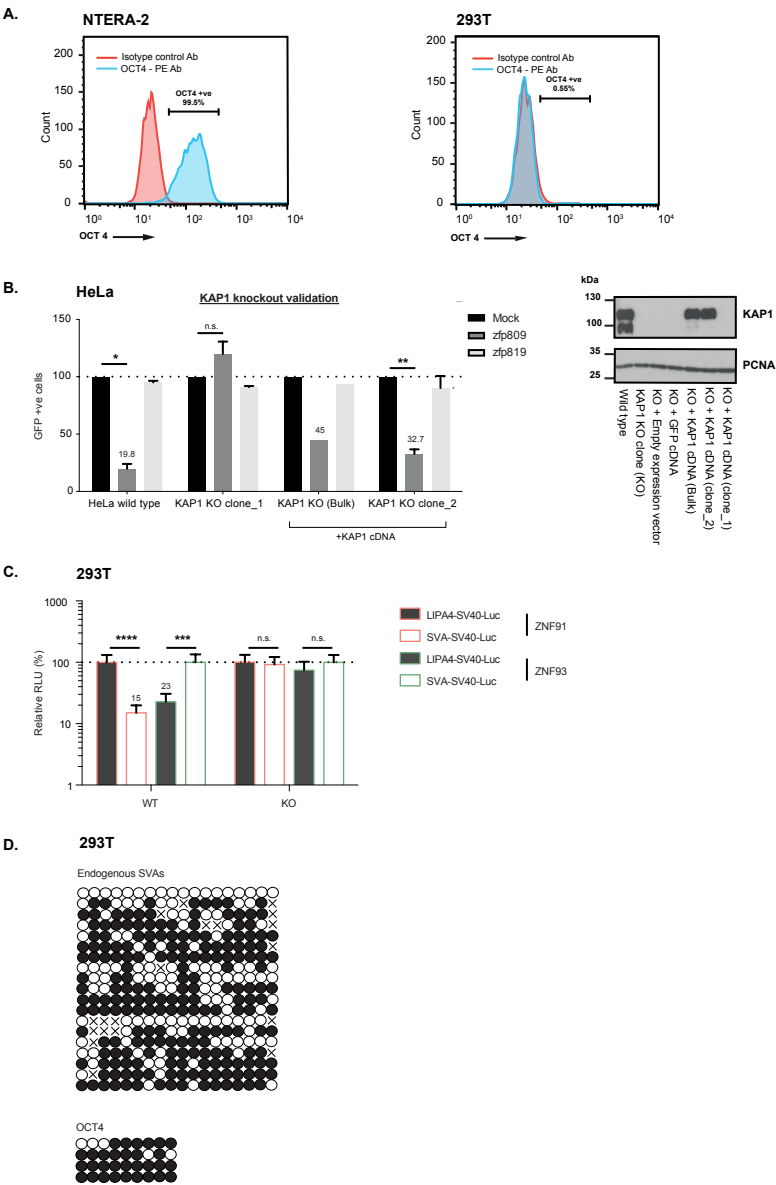


Figure S2

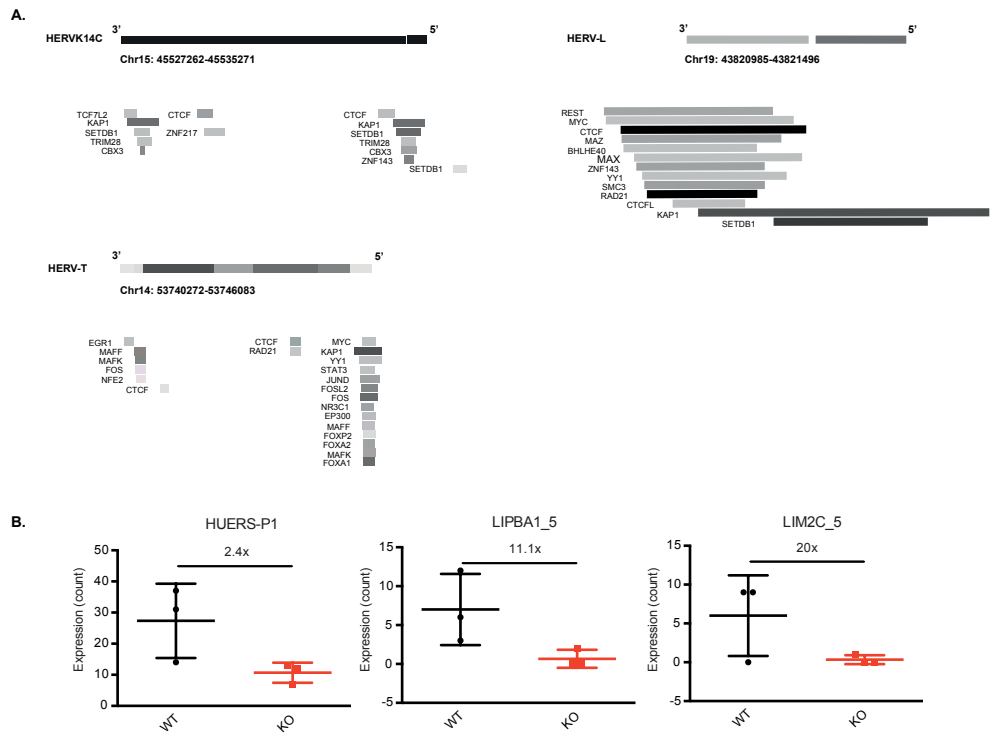


Figure S3

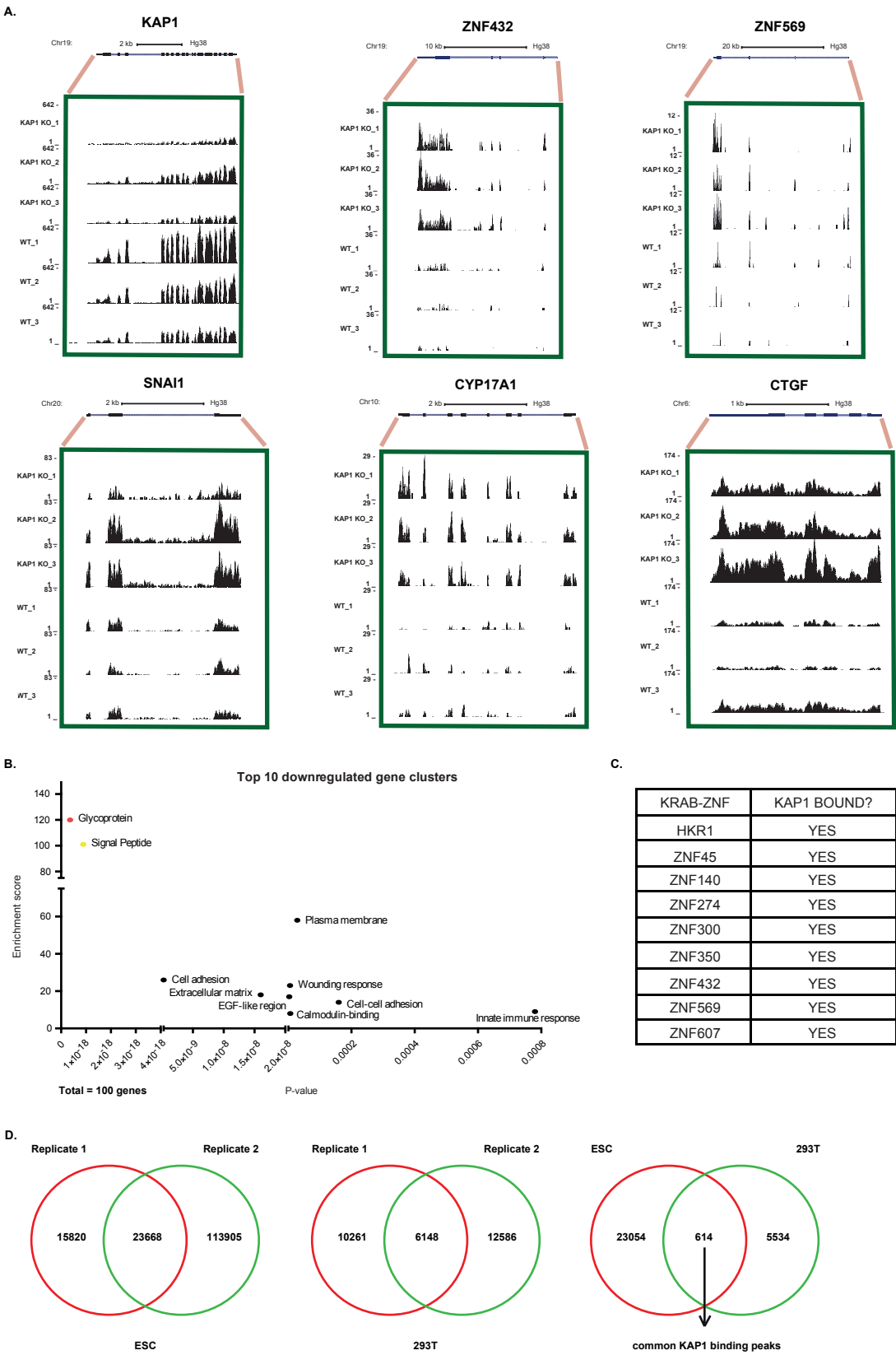


Figure S4

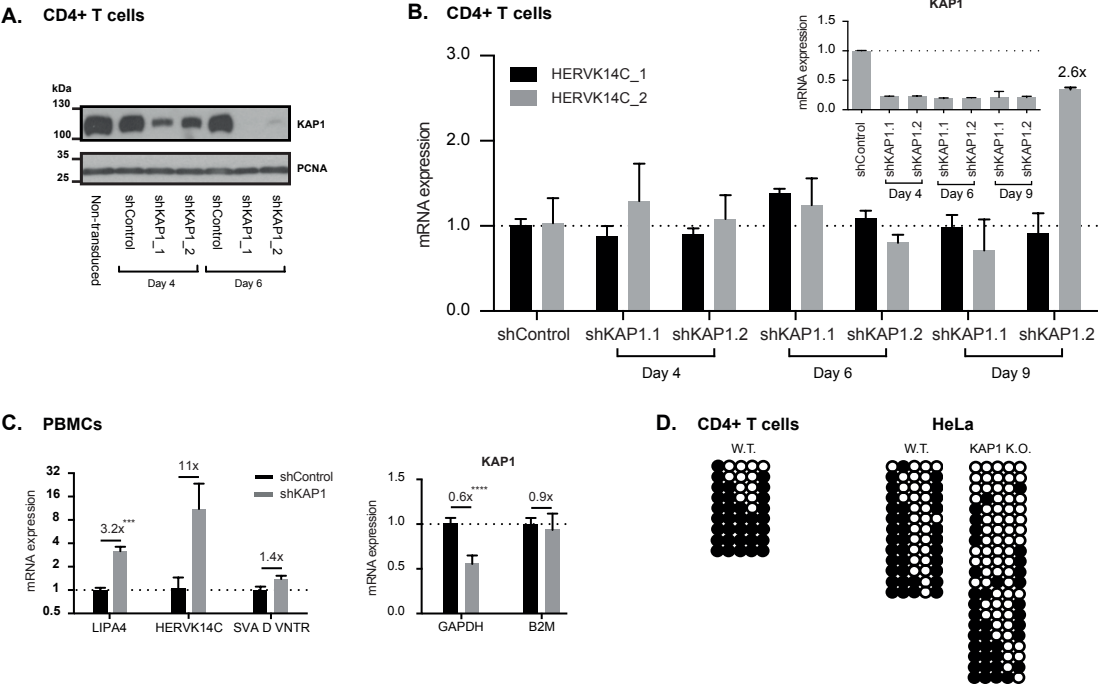
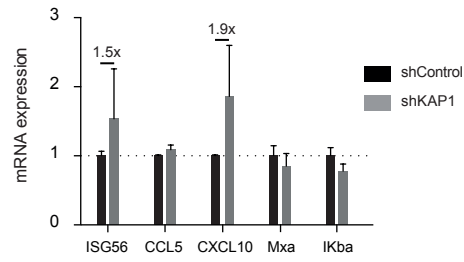
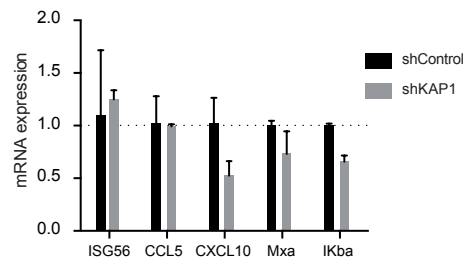


Figure S5

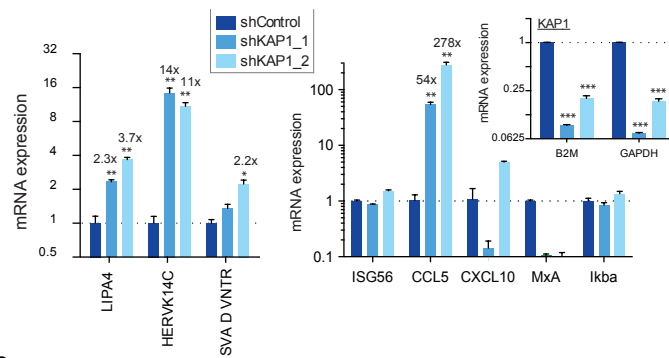
A. PBMCs



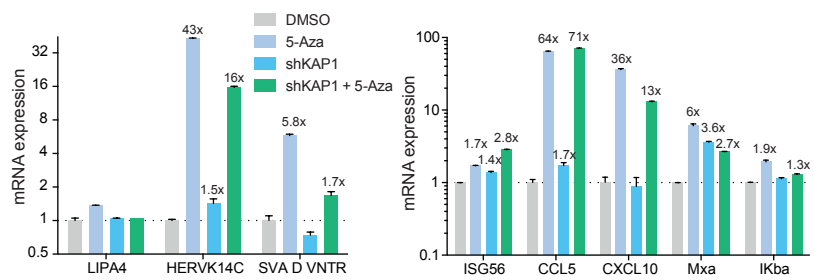
B. PBMCs



C. HeLa



C. 293T



“Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!”

— Lewis Carroll, *Alice Through the Looking Glass*